

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 66.1253 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-15

Perfect score: 34
Sequence: 1 gactcctcactatcgcagtcgctgagaca 34

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	34	21	AAA96849
2	19.2	56.5	543	22	AAH65584
3	19.2	56.5	612	22	AAH65585
4	19.2	56.5	666	22	AAF72234
5	19.2	56.5	349980	22	AAH68525
6	19	55.9	447	24	ABL89671
7	19	55.9	521	22	AAH41599
8	19	55.9	521	22	AAH29698
9	19	55.9	2874	22	AAH29578

C	10	19	55.9	3469	22	AAH41061	C	10	19	55.9	3469	22	AAH41061	C	10	19	55.9	3469	22	AAH41061
C	11	18.8	55.3	1015	21	AAH13367	C	11	18.8	55.3	1015	21	AAH13367	C	11	18.8	55.3	1015	21	AAH13367
C	12	18.8	55.3	1884	22	AAH04180	C	12	18.8	55.3	1884	22	AAH04180	C	12	18.8	55.3	1884	22	AAH04180
C	13	18.6	54.7	6075	24	ABL33550	C	13	18.6	54.7	6075	24	ABL33550	C	13	18.6	54.7	6075	24	ABL33550
C	14	18.6	54.7	611590	21	AAH22303	C	14	18.6	54.7	611590	21	AAH22303	C	14	18.6	54.7	611590	21	AAH22303
C	15	18.4	54.1	86	22	AAH23463	C	15	18.4	54.1	86	22	AAH23463	C	15	18.4	54.1	86	22	AAH23463
C	16	18.4	54.1	2601	23	AAH86004	C	16	18.4	54.1	2601	23	AAH86004	C	16	18.4	54.1	2601	23	AAH86004
C	17	18.4	54.1	3188	23	AAH8724	C	17	18.4	54.1	3188	23	AAH8724	C	17	18.4	54.1	3188	23	AAH8724
C	18	18.4	54.1	3336	24	AAH71356	C	18	18.4	54.1	3336	24	AAH71356	C	18	18.4	54.1	3336	24	AAH71356
C	19	18.4	54.1	6987	24	AAH53954	C	19	18.4	54.1	6987	24	AAH53954	C	19	18.4	54.1	6987	24	AAH53954
C	20	18.4	54.1	6990	24	AAH53959	C	20	18.4	54.1	6990	24	AAH53959	C	20	18.4	54.1	6990	24	AAH53959
C	21	18.4	54.1	1163020	24	AAH67197	C	21	18.4	54.1	1163020	24	AAH67197	C	21	18.4	54.1	1163020	24	AAH67197
C	22	18.4	54.1	3011208	24	AAH69245	C	22	18.4	54.1	3011208	24	AAH69245	C	22	18.4	54.1	3011208	24	AAH69245
C	23	18.2	53.5	884	18	AAH59476	C	23	18.2	53.5	884	18	AAH59476	C	23	18.2	53.5	884	18	AAH59476
C	24	18.2	53.5	11601	12	AAH013608	C	24	18.2	53.5	11601	12	AAH013608	C	24	18.2	53.5	11601	12	AAH013608
C	25	18.2	53.5	13058	14	AAH048231	C	25	18.2	53.5	13058	14	AAH048231	C	25	18.2	53.5	13058	14	AAH048231
C	26	18.2	53.5	30032	22	AAH17086	C	26	18.2	53.5	30032	22	AAH17086	C	26	18.2	53.5	30032	22	AAH17086
C	27	18.2	53.5	61890	23	AAH16586	C	27	18.2	53.5	61890	23	AAH16586	C	27	18.2	53.5	61890	23	AAH16586
C	28	18.2	53.5	534720	19	AAH30458	C	28	18.2	53.5	534720	19	AAH30458	C	28	18.2	53.5	534720	19	AAH30458
C	29	18.2	53.5	536165	19	AAH30459	C	29	18.2	53.5	536165	19	AAH30459	C	29	18.2	53.5	536165	19	AAH30459
C	30	18.2	52.9	565	23	AAH55233	C	30	18.2	52.9	565	23	AAH55233	C	30	18.2	52.9	565	23	AAH55233
C	31	18	52.9	4259	21	AAH74870	C	31	18	52.9	4259	21	AAH74870	C	31	18	52.9	4259	21	AAH74870
C	32	18	52.9	5305	20	AAH28289	C	32	18	52.9	5305	20	AAH28289	C	32	18	52.9	5305	20	AAH28289
C	33	18	52.9	5591	23	AAH04293	C	33	18	52.9	5591	23	AAH04293	C	33	18	52.9	5591	23	AAH04293
C	34	18	52.9	13437	23	AAH04292	C	34	18	52.9	13437	23	AAH04292	C	34	18	52.9	13437	23	AAH04292
C	35	18	52.9	1038602	20	AAH01425	C	35	18	52.9	1038602	20	AAH01425	C	35	18	52.9	1038602	20	AAH01425
C	36	18	52.9	1503900	22	AAH95240	C	36	18	52.9	1503900	22	AAH95240	C	36	18	52.9	1503900	22	AAH95240
C	37	18	52.9	1503900	22	AAH95733	C	37	18	52.9	1503900	22	AAH95733	C	37	18	52.9	1503900	22	AAH95733
C	38	18	52.4	371	23	AAH65536	C	38	18	52.4	371	23	AAH65536	C	38	18	52.4	371	23	AAH65536
C	39	17.8	52.4	1804	22	AAH09226	C	39	17.8	52.4	1804	22	AAH09226	C	39	17.8	52.4	1804	22	AAH09226
C	40	17.8	52.4	1804	22	AAH7850	C	40	17.8	52.4	1804	22	AAH7850	C	40	17.8	52.4	1804	22	AAH7850
C	41	17.8	52.4	1804	22	AAH85124	C	41	17.8	52.4	1804	22	AAH85124	C	41	17.8	52.4	1804	22	AAH85124
C	42	17.8	52.4	3371	22	AAH21261	C	42	17.8	52.4	3371	22	AAH21261	C	42	17.8	52.4	3371	22	AAH21261
C	43	17.8	52.4	1082138	21	AAH22305	C	43	17.8	52.4	1082138	21	AAH22305	C	43	17.8	52.4	1082138	21	AAH22305
C	44	17.6	51.8	507	24	AAH30419	C	44	17.6	51.8	507	24	AAH30419	C	44	17.6	51.8	507	24	AAH30419
C	45	17.6	51.8	507	24	AAH30419	C	45	17.6	51.8	507	24	AAH30419	C	45	17.6	51.8	507	24	AAH30419

ALIGNMENTS

RESULT 1	AAA96849	standard; DNA; 34 BP.
AAA96849	AAA96849	
AC	AAA96849	
AC	AAA96849	
DT	19-FEB-2001	(first entry)
XX		
DE	Guide desoxynucleotide building block G1.	
XX		
KW	Promoter; intergenic region; Comelina yellow mottle virus;	
KW	chimeric expression promoter; plant vascular expression promoter;	
KW	plant green tissue expression promoter; Cassava vein mosaic virus;	
KW	transgenic plant; ss.	
XX		
OS	Synthetic.	
XX		
PN	WO200058485-A1.	
XX		
PD	05-OCT-2000.	
XX		
PF	29-MAR-2000; 2000MO-IB00370.	
XX		
PR	29-MAR-1999; 99FR-0003925.	
XX		
XX	(MERI-) MERISTEM THERAPEUTICS.	
XX		
PI	Rance I, Gruber V, Theisen M;	
XX		
DR	cDNA encoding nove	
XX	WPI; 2000-647238/62.	
PT	Chimeric expression promoter for transgenic plant production, comprises	

PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Disclosure, Page 24, 91pp; English.
XX
CC The present sequence represents a guide desoxynucleotide building
CC block, which was used to link directional desoxynucleotide building
CC block during construction of chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Casseava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
SQ Sequence 34 BP; 8 A; 9 C; 7 G; 10 T; 0 other;
OY Query Match 100.0%; Score 34; DB 21; Length 34;
Best Local Similarity 100.0%; Pred. No. 3.5e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1 GACTCCTCTACTTATCGATCGTACTGTGAGACA 34
1 GACTCCTCTACTTATCGATCGTACTGTGAGACA 34
RESULT 2
AAH65584
ID AAH65584 standard; DNA; 543 BP.
AC AAH65584;
XX
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 619.
XX
XX
KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG90365.
XX
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX
PS Claim 8; SEQ ID NO: 619; 246BP + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from Corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
SQ Sequence 543 BP; 112 A; 160 C; 145 G; 126 T; 0 other;
OY Query Match 56.5%; Score 19.2; DB 22; Length 543;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
DB 1 GACTCCTCTACTTATCGATCGTACTGTGAGA 32
441 GACTCCTCTACTTATCGATCGTACTGTGAGA 472
RESULT 3
AAH65585/c
ID AAH65585 standard; DNA; 612 BP.
AC AAH65585;
XX
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 620.
XX
XX
KM Corynebacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EPI108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG90366.
XX
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX
PS Claim 8; SEQ ID NO: 620; 246BP + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from Corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

SO Sequence 612 BP, 147 A; 159 C; 171 G; 135 T; 0 other;

Query Match 56.5%; Score 19.2; DB 22; Length 612;
 Best Local Similarity 75.0%; Pred. No. 39;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGCTACTGTGAGA 32
 Db 154 GACTGCTCCTCTATCATCATCACTACGCTGACA 123

RESULT 4
 AAF72234
 ID AAF72234 standard; DNA; 666 BP.

AC AAF72234;

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:963.

XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;
 KM fine chemical production; microorganism; organic acid; nucleoside;
 KM nonproteogenic amino acid; purine base; pyrimidine base; nucleotide;
 KM lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KM carbohydrate; aromatic compound; cofactor; polypeptide; enzyme; ds.

XX Corynebacterium glutamicum.

PN WO200100843-A2.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000MO-IB00923.

XX 25-JUN-1999; 99US-0141031.
 PR 01-JUL-1999; 99DE-1030476.
 PR 02-JUL-1999; 99US-0142101.
 PR 08-JUL-1999; 99DE-1031415.
 PR 08-JUL-1999; 99DE-1031418.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.

PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041384.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.

PA (BADI) BASF AG.

PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Habernauer G;

DR WPI; 2001-137957/14.

XX P-PADB; AAB80115.

PT Nucleic acids from Corynebacterium glutamicum encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteogenic amino acids,
 PT and purine and pyrimidine bases -
 XX Claim 3; Page 1478; 1737pp; English.

XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals.
 CC in microorganisms, including organic acids, nonproteogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 CC compounds, vitamins, cofactors, polypeptides and enzymes.

XX Sequence 666 BP; 147 A; 189 C; 171 G; 159 T; 0 other;

Query Match 56.5%; Score 19.2; DB 22; Length 666;
 Best Local Similarity 75.0%; Pred. No. 40;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGCTACTGTGAGA 32
 Db 541 GACTGCTCCTCTATCATCATCACTACGCTGACA 572

RESULT 5

ID AAB68525

AC AAB68525;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 7060.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis; ds.
 XX Corynebacterium glutamicum.

PN EP108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOMA HAKKO KOGYO KK.
PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,
PI Tateishi N, Senoh A, Ikeda M, Ozaki A,
XX
XX WPI, 2001-376931/40.
DR
XX
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX
PS Disclosure; SEQ ID NO: 7060; 246bp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 349980 BP; 80289 A; 91081 C; 97378 G; 81232 T; 0 other;
XX
Query Match 56.5%; Score 19.2; DB 2; Length 349980;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
XX
QY 1 GACTCCTACTTATCGATCGTACTGTGAGA 32
DB 271214 GACTGCTCTCTGATCGATCGTACTGTGAGA 271245
XX
RESULT 6
ABL89671/c
ID ABL89671 standard; CDNA; 447 BP.
XX
AC ABL89671;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 233.
XX
XX Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnereary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200190304-A2.
XX
XX 29-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-US16450.
XX

PR 19-MAY-2000; 2000US-205515P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Birse CE, Rosen CA;
XX
XX WPI, 2002-122018/16.
DR P-FSDB; ABB89262.
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
PS Claim 4; SEQ ID NO 233; 2081bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 447 BP; 105 A; 120 C; 120 G; 101 T; 1 other;
XX
Query Match 55.9%; Score 19; DB 24; Length 447;
Best Local Similarity 81.5%; Pred. No. 46;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX
QY 1 GACTCCTACTTATCGATCGTACTG 27
DB 173 GACTGCGCTGTATGATCGTACTG 147
XX
RESULT 7
AAS41599/c
ID AAS41599 standard; CDNA; 521 BP.
XX
AC AAS41599;
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #815.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytosstatic;
KW anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS Homo sapiens.
XX
XX WO200155301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
XX

PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-MAR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PT Rosen CA, Barash SC, Ruben SM;
XX
XX
DR WPI; 2001-465566/50.
XX
XX P-PSDB; AAU23729.
XX
XX Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases

XX	PS	Claim 4; SEQ ID No 825; 1180bp; English.
XX	CC	The present invention relates to the isolation of novel human enzyme
XX	CC	polypeptides (AAU22915-AAU231814), and the cDNA and genomic sequences
XX	CC	encoding them. The enzyme polypeptides of the invention may comprise the
XX	CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,
XX	CC	isomerases or ligases. The sequences of the invention are useful in the
XX	CC	diagnosis, treatment, prevention and/or prognosis of a wide range of
XX	CC	disorders including hyperproliferative disorders (e.g. cancer),
XX	CC	immunodeficiency disorders (e.g. AIDS), autoimmune disorders
XX	CC	(e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
XX	CC	metabolic disorders (e.g. phenylketonuria), inflammatory disorders
XX	CC	(e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
XX	CC	blood-related disorders (e.g. haemophilia), reproductive disorders
XX	CC	(e.g. infertility) and infectious disorders (e.g. influenza). The
XX	CC	polynucleotides of the invention can also be used in gene therapy.
XX	CC	AA540785-AA541684 represent cDNA sequences encoding for the novel human
XX	CC	enzyme polypeptides of the invention.
XX	CC	Note: The sequence data for this patent did not form part of the printed
XX	CC	specification, but was obtained in electronic format directly from Wiro
XX	CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	SO	Sequence 521 BP, 120 A, 145 C, 135 G, 116 T, 5 other;
OY	Query Match	55.9%; Score 19; DB 22; Length 521;
Db	Best Local Similarity	81.5%; Pred. No. 48;
	Matches 22; Conservative	0; Mismatches 5; Indels 0; Gaps 0;
	1 GACTCCTCACTTATGATCGGTACTG 27	
	198 GACTGCCTGCTTATGATCGGTACTG 172	
RESULT 8		
ID	AA529698/C	
XX	AA529698 standard; cDNA, 521 BP.	
AC		
XX	AA529698;	
DT	21-NOV-2001 (first entry)	
XX		
DE	Human endocrine polypeptide encoding cDNA SEQ ID No 198.	
XX		
KW	Endocrine protein; human; mouse; rabbit; goat; horse; food additive;	
KW	cac; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;	
KW	antirheumatic; antiproliferative; cytosolic; cardiac; neuroprotective;	
KW	cerebroprotective; nootropic; antibacterial; vitinoid; fungicide; cancer;	
KW	ophthalmological; vulnery; gene therapy; autoimmune diseases; neoplasm;	
KW	hyperproliferative disorder; breast; liver; cardiovascular disorder; ss;	
KW	cerebrovascular disorder; nervous system disorder; bacterial infection;	
KW	fungal infection; viral infection; ocular disorder; endocrine disorder;	
KW	gastrointestinal disorder; renal disorder; respiratory disorder;	
KW	wound healing; skin aging; organ transplantation; food preservative;	
KW	tissue regeneration; anti-infertility.	
OS	Homo sapiens.	
XX		
PN	WO200155364-A2.	
PD		
XX	02-AUG-2001.	
PF		
XX	17-JAN-2001; 2001WO-US01308.	
XX		
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	

CC	20-OCT-2000;	2000US-0241786.
CC	20-OCT-2000;	2000US-0241787.
CC	20-OCT-2000;	2000US-0241808.
CC	20-OCT-2000;	2000US-0241809.
CC	20-OCT-2000;	2000US-0241826.
CC	01-NOV-2000;	2000US-0244617.
CC	08-NOV-2000;	2000US-0246474.
CC	08-NOV-2000;	2000US-0246475.
CC	08-NOV-2000;	2000US-0246476.
CC	08-NOV-2000;	2000US-0246477.
CC	08-NOV-2000;	2000US-0246478.
CC	08-NOV-2000;	2000US-0246523.
CC	08-NOV-2000;	2000US-0246524.
CC	08-NOV-2000;	2000US-0246525.
CC	08-NOV-2000;	2000US-0246526.
CC	08-NOV-2000;	2000US-0246527.
CC	08-NOV-2000;	2000US-0246528.
CC	08-NOV-2000;	2000US-0246532.
CC	08-NOV-2000;	2000US-0246609.
CC	08-NOV-2000;	2000US-0246610.
CC	08-NOV-2000;	2000US-0246611.
CC	08-NOV-2000;	2000US-0246613.
CC	17-NOV-2000;	2000US-0249207.
CC	17-NOV-2000;	2000US-0249208.
CC	17-NOV-2000;	2000US-0249209.
CC	17-NOV-2000;	2000US-0249210.
CC	17-NOV-2000;	2000US-0249211.
CC	17-NOV-2000;	2000US-0249212.
CC	17-NOV-2000;	2000US-0249213.
CC	17-NOV-2000;	2000US-0249214.
CC	17-NOV-2000;	2000US-0249215.
CC	17-NOV-2000;	2000US-0249216.
CC	17-NOV-2000;	2000US-0249217.
CC	17-NOV-2000;	2000US-0249218.
CC	17-NOV-2000;	2000US-0249244.
CC	17-NOV-2000;	2000US-0249245.
CC	17-NOV-2000;	2000US-0249264.
CC	17-NOV-2000;	2000US-0249265.
CC	17-NOV-2000;	2000US-0249297.
CC	17-NOV-2000;	2000US-0249329.
CC	17-NOV-2000;	2000US-0249330.
CC	01-DEC-2000;	2000US-0250160.
CC	01-DEC-2000;	2000US-0250391.
CC	05-DEC-2000;	2000US-0251030.
CC	05-DEC-2000;	2000US-0251988.
CC	05-DEC-2000;	2000US-0256719.
CC	06-DEC-2000;	2000US-0251479.
CC	08-DEC-2000;	2000US-0251856.
CC	08-DEC-2000;	2000US-0251868.
CC	08-DEC-2000;	2000US-0251869.
CC	08-DEC-2000;	2000US-0251989.
CC	08-DEC-2000;	2000US-0251997.
CC	11-DEC-2000;	2000US-0254097.
CC	05-JAN-2001;	2001US-0259678.
CC	(HUMA-) HUMAN GENOME SCI INC.	
CC	Rosen CA, Barash SC, Ruben SW;	
CC	WPI, 2001-451936/48.	
CC	P-PSDB; AAUI8469.	
CC	Isolated polypeptide for treating, preventing and/ or prognosing	
CC	disorders of the endocrine system such as reproductive disorders,	
CC	endocrine cancers and also for testing and detection e.g. diagnosis -	
CC	claim 1, SEQ ID No 198; 604bp; English.	
CC	Sequences AAS29511-AAS29736 represent cDNA molecules, which encode the	
CC	endocrine polypeptides of the invention. Endocrine polypeptides and their	
CC	associated polynucleotides of the invention are useful in the diagnosis,	
CC	treatment and prevention of various types of disorders in e.g. humans,	
CC	mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A	

CC	pathological condition can be determined by determining the presence or absence of a mutation in an endocrine polynucleotide. The treatable disorders include autoimmune diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischaemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi, ocular disorders such as corneal infection, endocrine disorders such as premature labour and infertility, gastrointestinal disorders such as Crohn's disease, renal disorders such as glomerulonephritis and respiratory disorders such as asthma. The polypeptides can also be used to aid wound healing, to prevent skin aging due to sunburn, to maintain organs before transplantation, to regenerate tissues and in chemoraxs. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities.
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO
Query Match	55.9%; Score 19; DB 22; Length 521;
Best Local Similarity	81.5%; Pred. No. 48;
Matches	22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy	1 GACTCCTCTACTTATTCAGTCGGTACTG 27
Db	198 GACTCGCGTCTGTTATGATCGTACTG 172
RESULT 9	
AA529578/c	
ID	AA529578 standard; cDNA; 2874 BP.
XX	AA529578;
AC	
XX	
DT	21-NOV-2001 (first entry)
XX	
DE	Human endocrine polypeptide encoding cDNA SEQ ID No 78.
XX	
KW	Endocrine protein; human; mouse; rabbit; goat; horse; food additive; cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic; antineumatic; antiproliferative; cyostatic; cardiant; neuroprotective; cerebroprotective; nocrotropic; antibacterial; vincinide; fungicide; cancer; ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm; hyperproliferative disorder; breast; liver; cardiovascular disorder; ss; cerebrovascular disorder; nervous system disorder; bacterial infection; fungal infection; viral infection; ocular disorder; endocrine disorder; gastrointestinal disorder; renal disorder; respiratory disorder;
KW	wound healing; skin aging; organ transplantation; food preservative; tissue regeneration; anti-sterility.
XX	
OS	Homo sapiens.
XX	
XX	WO20015364-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01308.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.

PR	6-5-UTL-2000	2000US-02209664
PR	26-4-UTL-2000	2000US-02209664
PR	14-AUG-2000	2000US-02245518
PR	14-AUG-2000	2000US-02245519
PR	14-AUG-2000	2000US-02245519
PR	14-AUG-2000	2000US-02245513
PR	14-AUG-2000	2000US-02245513
PR	14-AUG-2000	2000US-02245514
PR	14-AUG-2000	2000US-02252665
PR	14-AUG-2000	2000US-02252667
PR	14-AUG-2000	2000US-02252668
PR	18-AUG-2000	2000US-02262759
PR	22-AUG-2000	2000US-02262681
PR	22-AUG-2000	2000US-02262681
PR	22-AUG-2000	2000US-02271882
PR	23-AUG-2000	2000US-02271882
PR	30-AUG-2000	2000US-02289294
PR	01-SEP-2000	2000US-02282987
PR	01-SEP-2000	2000US-02293343
PR	01-SEP-2000	2000US-02293344
PR	01-SEP-2000	2000US-02293345
PR	05-SEP-2000	2000US-02285013
PR	05-SEP-2000	2000US-02285013
PR	06-SEP-2000	2000US-02304337
PR	06-SEP-2000	2000US-02304338
PR	06-SEP-2000	2000US-02312422
PR	08-SEP-2000	2000US-02312443
PR	08-SEP-2000	2000US-02312443
PR	08-SEP-2000	2000US-02314413
PR	08-SEP-2000	2000US-02314414
PR	08-SEP-2000	2000US-02330804
PR	08-SEP-2000	2000US-02330811
PR	12-SEP-2000	2000US-02319681
PR	14-SEP-2000	2000US-02323297
PR	14-SEP-2000	2000US-02333998
PR	14-SEP-2000	2000US-02333998
PR	14-SEP-2000	2000US-02343401
PR	14-SEP-2000	2000US-02343401
PR	14-SEP-2000	2000US-02330633
PR	14-SEP-2000	2000US-02330634
PR	14-SEP-2000	2000US-02330655
PR	21-SEP-2000	2000US-02342223
PR	21-SEP-2000	2000US-02342223
PR	25-SEP-2000	2000US-02349997
PR	25-SEP-2000	2000US-02349998
PR	26-SEP-2000	2000US-02349984
PR	27-SEP-2000	2000US-02358344
PR	27-SEP-2000	2000US-02358344
PR	29-SEP-2000	2000US-02363277
PR	29-SEP-2000	2000US-02363277
PR	29-SEP-2000	2000US-02363657
PR	29-SEP-2000	2000US-02363658
PR	29-SEP-2000	2000US-02363659
PR	29-SEP-2000	2000US-02363670
PR	02-OCT-2000	2000US-02368022
PR	02-OCT-2000	2000US-02370337
PR	02-OCT-2000	2000US-02370338
PR	02-OCT-2000	2000US-02370339
PR	02-OCT-2000	2000US-02370339
PR	02-OCT-2000	2000US-02370440
PR	13-OCT-2000	2000US-02393935
PR	13-OCT-2000	2000US-02393937
PR	20-OCT-2000	2000US-02409600
PR	20-OCT-2000	2000US-02412221
PR	20-OCT-2000	2000US-02417885
PR	20-OCT-2000	2000US-02417885
PR	20-OCT-2000	2000US-02417876
PR	20-OCT-2000	2000US-02418809
PR	20-OCT-2000	2000US-02418809
PR	20-OCT-2000	2000US-02418826
PR	01-NOV-2000	2000US-02464474
PR	08-NOV-2000	2000US-02464474

PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246479
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246539
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249308
PR	17-NOV-2000	2000US-0249309
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250391
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251088
PR	05-DEC-2000	2000US-0251671
PR	06-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251868
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251893
PR	11-DEC-2000	2000US-0251907
PR	05-JAN-2001	2001US-0259678
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Barash SC, Ruben SM,	
DR	WPI, 2001-451936/48.	
XX	P-PSDB; ANU18345.	
XX	Isolated polypeptide for treatin	
XX	disorders of the endocrine syste	
XX	endocrine cancers and also for t	
XX	Claim 1; SEQ ID No 78; 604pp; E	
XX	Sequences AA529511-AA529736 rep	
XX	endocrine polypeptides of the in	
XX	associated polynucleotides of th	
XX	treatment and prevention of vari	
XX	mice, rabbits, goats, horses, ca	
XX	pathological condition can be det	
XX	absence of a mutation in an endo	
XX	disorders include autoimmune dis	
XX	hyperproliferative disorders suc	
XX	cardiovascular disorders such as	
XX	disorders such as cerebral ischa	
XX	Alzheimer's disease, infections	

Isolated polypeptide for treating, preventing and/or prognosing disorders of the endocrine system such as reproductive disorders, endocrine cancers and also for testing and detection e.g. diagnosis -

Claim 1; SEQ ID No 78; 604bp; English.

Sequences AAS29511-AAS29736 represent cDNA molecules, which encode the associated polynucleotides of the invention. Endocrine polypeptides and their associated polynucleotides of the invention are useful in the diagnosis, treatment and prevention of various types of disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A pathological condition can be determined by determining the presence or absence of a mutation in an endocrine polynucleotide. The treatable disorders include autoimmune diseases such as rheumatoid arthritis, hyperproliferative disorders such as neoplasms of the breast or liver, cardiovascular disorders such as cardiac arrest, cerebrovascular disorders such as cerebral ischaemia, nervous system disorders such as Alzheimer's disease, infections caused by bacteria, viruses and fungi.

CC ocular disorders such as corneal infection, endocrine disorders such as
CC premature labour and infertility, gastrointestinal disorders such as
CC Crohn's disease, renal disorders such as glomerulonephritis and
CC respiratory disorders such as asthma. The polypeptides can also be used
CC to aid wound healing, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, to regenerate tissues and in chemotaxis.
CC The polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

Query Match 55.9%; Score 19; DB 22; Length 2874;
Best Local Similarity 81.5%; Pred. No. 62;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 GACTGCTTACTTATCGATCGTACTG 27
Db 198 GACTGCTTACTTATCGATCGTACTG 172

RESULT 10

AA541061/c
ID AA541061 standard; cDNA; 3489 BP.

XX AA541061;

DT 17-DEC-2001 (first entry)

XX cDNA encoding novel human enzyme polypeptide #277.

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
XX ligase; hyperproliferative disorder; immunodeficiency disorder;
XX autoimmune disorder; neurological disorder; metabolic disorder;
XX inflammatory disorder; cardiovascular disorder; reproductive disorder;
XX blood-related disorder; infectious disorder; gene therapy; cytostatic;
XX anti arthritic; nephrotropic; anticoagulant; ss.

OS Homo sapiens.

XX W0200155301-A2.

PD 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01239.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 11-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232402.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 25-SEP-2000; 2000US-0234599.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.

PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249224.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-), HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465566/50.
 DR P-PSDB; AAU23191.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 XX
 PS Claim 4; SEQ ID NO 287; 1180bp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AA840785-AA841684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 3489 BP; 835 A; 859 C; 960 G; 832 T; 3 other;

Query Match 55.9%; Score 19; DB 22; Length 3489;

Best Local Similarity 81.5%; Pred. No. 64;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 GACTCCTCTACTTATGATCGTACTG 27
 Db 212 GACTGCGCTGCTTATGATCGTACTG 186
 RESULT 11
 AAF13367
 ID AAF13367 standard; cDNA; 1015 BP.
 XX
 AC AAF13367;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:5890.
 XX
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 PN WO200056762-A2.
 XX
 FD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000MO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 DR WPI; 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 88; Page 2434; 3161bp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organization of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 CC
 XX
 SQ Sequence 1015 BP; 233 A; 280 C; 259 G; 243 T; 0 other;

Query Match 55.3%; Score 18.8; DB 21; Length 1015;
 Best Local Similarity 76.7%; Pred. No. 65;
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ACTCCTCTACTTATGATCGGCTACTGTGAG 31
 DB 439 AGTCCTCTACTTATGATCGGCTACTGTG 468

RESULT 12

ADD04180
 ID AAD04180 standard; DNA; 1884 BP.

XX AAD04180;

DT 02-JUL-2001 (first entry)

DE Aspergillus oryzae lysophospholipase 1 (LLPL-1) DNA.

XX Lysophospholipase; LLPL-1; lysophospholipid; lecithin; filterability;

KM bakery product; dough; bread; cake; elasticity; edible oil; rape seed;

XX soybean; sunflower; ds.

XX Aspergillus oryzae.

XX Key Location/Qualifiers

FT CDS 1..1884

FT /tag= a

FT /product= "Aspergillus oryzae LLPL-1 protein"

FT sig_peptide 1..63

FT /tag= b

FT mat_peptide 64..1881

FT /tag= c

FT /product= "Mature A. oryzae LLPL-1 protein"

XX WO200127251-A1.

XX 19-APR-2001.

XX 12-OCT-2000; 2000WO-DK00577.

XX 14-OCT-1999; 99DK-0001473.

XX (NOVO) NOVOTYMES AS.

XX Udagawa H, Frandsen TP, Nielsen TAB, Kaupinen MS, Christensen S;

XX WPI; 2001-282018/29.

XX P-PSDB; AAE00814.

XX New lysophospholipase enzyme useful in hydrolyzing fatty acyl groups in

XX a phospholipid or lysophospholipid, and in improving filterability of

XX an aqueous solution or slurry of carbohydrate origin -

XX Claim 4; Page 39-42; 58pp; English.

XX The present sequence is a DNA encoding Aspergillus oryzae

XX lysophospholipase-1 (LLPL-1). Lysophospholipase is an enzyme that

XX hydrolyses lysophospholipids into fatty acids. The DNA encoding LLPL-1

XX is cloned into a plasmid present in Escherichia coli deposit number

XX DSM 13082. Lysophospholipase is used to hydrolyse fatty acyl groups

XX in a phospholipid or lysophospholipid and for improving filterability

XX of an aqueous solution or slurry of carbohydrate origin which contains

XX phospholipid and preferably a starch hydrolyase, particularly wheat

XX starch hydrolyase. Lysophospholipase is also used to release fatty

XX acids from intact phospholipid e.g. lecithin, and in the preparation

XX of bakery products such as dough, bread, and cakes to improve elasticity.

XX Lysophospholipase is used to reduce the phospholipid content in an

XX edible oil e.g. vegetable oils such as rape seed oil, soybean oil and

XX sunflower oil.

XX Sequence 1884 BP; 429 A; 556 C; 463 G; 436 T; 0 other;

Query Match 55.3%; Score 18.8; DB 22; Length 1884;
 Best Local Similarity 76.7%; Pred. No. 72;
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ACTCCTCTACTTATGATCGGCTACTGTGAG 31
 DB 425 AGTCCTCTACTTATGATCGGCTACTGTG 454

RESULT 13

ABL33550/c
 ID ABL33550 standard; DNA; 6075 BP.

XX ABL33550;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1523.

XX Human; immune system disease; cytosine methylation; antiaesthetic;

KM antineoplastic; anti-HIV; anticonvulsant; ophthalmological;

KM antineoplastic; antineoplastic; antidiabetic; antipsoriatic;

KM antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;

KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KM gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIDEMIOLOGICAL AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful

XX for diagnosis and treatment of diseases associated with abnormal

XX cytosine methylation -

XX Claim 1; SEQ ID NO 1523; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

XX genes which are modified by the methylation of cytosines. The sequences

XX can be used in the diagnosis and treatment of immune system disorders,

XX including eye diseases such as retinopathy, neovascular glaucoma and

XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

XX diseases. The present sequence is a gene of the invention.

XX Sequence 6075 BP; 1318 A; 146 C; 1645 G; 2966 T; 0 other;

XX Query Match 54.7%; Score 18.6; DB 24; Length 6075;

XX Best Local Similarity 72.7%; Pred. No. 1.1e+02;

XX Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 ACTCCTCTACTTATGATCGGCTACTGTGAGCA 34

DB 859 ACACCTCTACTTATGATCGGCTACTGTG 827

RESULT 14

```

AAF22303
ID AAF22303 standard; DNA; 611590 BP.
XX
AC AAF22303;
XX
DT 20-MAR-2001 (first entry)
XX
DE Arabidopsis thaliana chromosome 2 centromere.
XX
KM Centromere; microsome; vector; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200055325-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-US07392.
XX
PR 18-MAR-1999; 99US-0125219.
PR 01-APR-1999; 99US-0127409.
PR 18-MAY-1999; 99US-0134770.
PR 13-SEP-1999; 99US-0133584.
PR 17-SEP-1999; 99US-0154603.
XX
PA (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Keith K;
XX
DR WPI; 2000-587529/55.
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microsome which can serve as vectors for
PT the construction of transgenic plant and animal cells
XX
PS Claim 45; Page 820-959; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microsome which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 184983 T; 0 other;
Query Match 54.7%; Score 18.6; DB 21; Length 611590;
Best Local Similarity 72.7%; Pred. No. 2.1e+02;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 GACTCCTCTACTTATCGATCGGTACTGTGAGC 33
Db 384616 GAGACATCTACTTACTTATCGGTACTGTGAGC 384648

```

RESULT 15
 AAS23463/c
 ID AAS23463 standard; DNA; 86 BP.

```

XX
AC AAS23463;
XX
DT 04-DEC-2001 (first entry)
XX
DE C. albicans essential gene knockout upstream PCR primer (KO-Up) #21.
XX
KM Gene identification; essential gene; GRACE; pathogenic fungus;
KM gene replacement and conditional expression; fungal infection;
XX PCR primer; KO-Up; ss.
XX
OS Candida albicans.
XX
PN WO200160975-A2.
XX

```

```

PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05551.
XX
PR 18-FEB-2000; 2000US-0183534.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H;
XX
DR WPI; 2001-489080/53.
XX
PT Identifying genes essential to fungal metabolisms and identifying
PT potential therapeutic agents that target these genes
XX
PS Disclosure; Page 255; 324pp; English.
XX
CC The present invention relates to novel methods for constructing fungal
CC strains useful for identification and validation of gene products as
CC targets for therapeutic agents, for creating a collection of identified
CC essential genes, and screening assays for the discovery of new drugs.
CC The invention provides the GRACE (gene replacement and conditional
CC expression) method for the construction of mutant organisms referred to
CC as GRACE strains of the organism. The invention can be applied to any
CC organism, particularly a pathogenic fungus e.g. Candida albicans,
CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are
CC useful to identify agents that may be used in the treatment of fungal
CC infections. AAS23463-AAS23503 represent C. albicans essential gene
CC knockout upstream PCR primers (KO-Up) used in the methods of the
CC present invention.
XX
SQ Sequence 86 BP; 28 A; 20 C; 10 G; 28 T; 0 other;
Query Match 54.1%; Score 18.4; DB 22; Length 86;
Best Local Similarity 78.6%; Pred. No. 68;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 4 TCCCTACTTATCGATCGGTACTGTGAG 31
Db 83 TCCACTAGTTCTAGAGCGTTCTGTGAG 56

```

Search completed: May 11, 2003, 03:06:06
 Job time : 188.125 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 13.6341 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-15

Perfect score: 34

Sequence: 1 gactcctcattcgcacgctgagaca 34

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18.4	54.1	3336	4	US-09-026-033-1	Sequence 1, Appli
C 2	18.4	54.1	3336	4	US-09-026-033-2	Sequence 2, Appli
C 3	18.4	54.1	6987	4	US-09-026-033-3	Sequence 3, Appli
C 4	18.4	54.1	6990	4	US-09-026-033-23	Sequence 23, Appli
C 5	18.2	53.5	11601	2	US-08-222-617A-3	Sequence 24, Appli
C 6	18.2	53.5	11601	2	US-08-222-617A-24	Sequence 24, Appli
C 7	17.6	51.8	507	4	US-09-280-116-189	Sequence 189, App
C 8	17.6	51.8	933	4	US-09-280-116-191	Sequence 191, App
C 9	17.4	51.2	2432	3	US-08-974-022-1	Sequence 1, Appli
C 10	17.4	51.2	2432	4	US-08-795-445A-1	Sequence 1, Appli
C 11	17.4	51.2	2432	4	US-08-974-186-1	Sequence 1, Appli
C 12	17.4	51.2	2432	4	US-08-795-446B-1	Sequence 1, Appli
C 13	17.4	51.2	2432	4	US-08-795-446B-1	Sequence 1, Appli
C 14	17.4	51.2	2432	4	US-08-795-446B-1	Sequence 1, Appli
C 15	17.2	50.6	993	1	US-08-364-081-2	Sequence 2, Appli
C 16	17.2	50.6	993	1	US-08-630-552-2	Sequence 2, Appli
C 17	17.2	50.6	993	5	PCT-US95-16558-2	Sequence 2, Appli
C 18	17.2	50.6	1011	1	US-08-375-235-1	Sequence 1, Appli
C 19	17.2	50.6	1183	1	US-07-755-009-1	Sequence 1, Appli
C 20	17.2	50.6	1183	1	US-07-755-009-2	Sequence 1, Appli
C 21	17.2	50.6	1860	2	US-08-331-644-3	Sequence 3, Appli
C 22	17.2	50.6	1860	5	PCT-US93-04102-3	Sequence 3, Appli
C 23	17.2	50.6	4910	4	US-09-103-064-40	Sequence 40, Appli
C 24	17.2	50.6	4910	4	US-09-086-503-40	Sequence 40, Appli
C 25	17.2	50.0	249	4	US-09-134-001C-1324	Sequence 1324, Ap
C 26	17.2	50.0	1303	3	US-08-894-440-2	Sequence 2, Appli
C 27	17.2	50.0	1303	4	US-09-458-093-2	Sequence 2, Appli

C 28	17	50.0	17710	4	US-08-976-259-70	Sequence 70, Appli
C 29	16.8	49.4	240	1	US-08-054-985-1	Sequence 1, Appli
C 30	16.8	49.4	1177	3	US-08-611-587-7	Sequence 7, Appli
C 31	16.8	49.4	1240	4	US-09-221-017B-844	Sequence 844, App
C 32	16.6	48.8	500	2	US-08-967-101-50	Sequence 50, Appli
C 33	16.6	48.8	500	2	US-08-592-541-50	Sequence 50, Appli
C 34	16.6	48.8	500	3	US-09-124-698-50	Sequence 50, Appli
C 35	16.6	48.8	500	3	US-09-124-698-50	Sequence 50, Appli
C 36	16.6	48.8	500	4	US-08-496-841C-50	Sequence 50, Appli
C 37	16.6	48.8	500	4	US-09-124-523-50	Sequence 50, Appli
C 38	16.6	48.8	611	4	US-09-221-017B-352	Sequence 352, App
C 39	16.6	48.8	736	2	US-08-967-101-12	Sequence 12, Appli
C 40	16.6	48.8	736	2	US-08-967-101-155	Sequence 155, App
C 41	16.6	48.8	736	2	US-08-592-541-12	Sequence 12, Appli
C 42	16.6	48.8	736	2	US-08-592-541-155	Sequence 155, App
C 43	16.6	48.8	736	3	US-08-888-077A-15	Sequence 15, Appli
C 44	16.6	48.8	736	3	US-09-124-698-12	Sequence 12, Appli
C 45	16.6	48.8	736	3	US-09-124-698-155	Sequence 155, App

ALIGNMENTS

RESULT 1
US-09-026-033-1/c
Sequence 1, Application US/09026033
Patent No. 6368791
GENERAL INFORMATION:
APPLICANT: Felix, Carolyn A.
APPLICANT: Jones, Douglas H.
APPLICANT: Rappaport, Eric
TITLE OF INVENTION: METHOD AND KITS FOR ANALYSIS OF
TITLE OF INVENTION: CHROMOSOMAL REARRANGEMENTS ASSOCIATED WITH LEUKEMIA
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PANITCH SCHWARZ JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,033
FILING DATE: 19-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/038,624
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/056,923
FILING DATE: 26-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/065,911
FILING DATE: 17-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary, Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3336 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-026-033-1

Query Match 54.1%; Score 18.4; DB 4; Length 3336;
Best Local Similarity 78.6%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 7 TCTACTATGATCGGCTGCTGTGAGACA 34
Db 793 TTACTATTGATTGGCAGCTGTGGCCA 766

RESULT 2

US-09-026-033-2
Sequence 2, Application US/09026033
Patent No. 6368791
GENERAL INFORMATION:
APPLICANT: Felix, Carolyn A.
APPLICANT: Jones, Douglas H.
APPLICANT: Rappaport, Eric
TITLE OF INVENTION: METHOD AND KITS FOR ANALYSIS OF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
STREET: Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,033
FILING DATE: 19-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/038,624
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/056,923
FILING DATE: 26-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/065,911
FILING DATE: 17-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary, Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-1001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3336 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-026-033-2

Query Match 54.1%; Score 18.4; DB 4; Length 3336;
Best Local Similarity 78.6%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 7 TCTACTATGATCGGCTGCTGTGAGACA 34
Db 793 TTACTATTGATTGGCAGCTGTGGCCA 766

Db 2544 TTACTATTGATTGGCAGCTGTGGCCA 2571

RESULT 3

US-09-026-033-3
Sequence 3, Application US/09026033
Patent No. 6368791
GENERAL INFORMATION:
APPLICANT: Felix, Carolyn A.
APPLICANT: Jones, Douglas H.
APPLICANT: Rappaport, Eric
TITLE OF INVENTION: METHOD AND KITS FOR ANALYSIS OF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
STREET: Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,033
FILING DATE: 19-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/038,624
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/056,923
FILING DATE: 26-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/065,911
FILING DATE: 17-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary, Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-1001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6987 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-026-033-3

Query Match 54.1%; Score 18.4; DB 4; Length 6987;
Best Local Similarity 78.6%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 7 TCTACTATGATCGGCTGCTGTGAGACA 34
Db 6195 TTACTATTGATTGGCAGCTGTGGCCA 6222

RESULT 4

US-09-026-033-23
Sequence 23, Application US/09026033
Patent No. 6368791
GENERAL INFORMATION:
APPLICANT: Felix, Carolyn A.
APPLICANT: Jones, Douglas H.

APPLICANT: Rappaport, Eric
TITLE OF INVENTION: METHOD AND KITS FOR ANALYSIS OF
TITLE OF INVENTION: CHROMOSOMAL REARRANGEMENTS ASSOCIATED WITH LEUKEMIA
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
STREET: One Commerce Square, 2005 Market Street, 22nd
FLOOR
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,033
FILING DATE: 19-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/038,624
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/056,923
FILING DATE: 26-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/065,911
FILING DATE: 17-NOV-1997
ATTORNEY/AGENT INFORMATION:
NAME: Doyle Leary, Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 7600-10011
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 6990 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-026-033-23

Query Match 54.1%; Score 18.4; DB 4; Length 6990;
Best Local Similarity 78.6%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 TCTACTATGATCGGTAAGTGTGACA 34
DB 6196 TTACTATGATGATGCGACTGTGGCCA 6223

RESULT 5
US-08-222-617A-3
Sequence 3, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan P.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase

NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-Apr-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 11601 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Acremonium chrysogenum
FEATURE:
NAME/KEY: CDS
LOCATION: 8050..8052
OTHER INFORMATION: /function= "Enzyme"
OTHER INFORMATION: /product= "ACV Synthetase"
OTHER INFORMATION: /note= "NNN=GCC, AGU, AGC, UCU, UCC, UCA, or UCG;"
US-08-222-617A-3

Query Match 53.5%; Score 18.2; DB 2; Length 11601;
Best Local Similarity 87.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 10 ACTTATGATCGGTAAGTGTGACA 32
DB 11163 ACTTATGATCGGTAAGTGTGACA 11185

RESULT 6
US-08-222-617A-24
Sequence 24, Application US/08222617A
Patent No. 5882879
GENERAL INFORMATION:
APPLICANT: Veenstra, Annemarie E.
APPLICANT: Martin, Juan P.
APPLICANT: Garcia, Bruno D.
APPLICANT: Gutierrez, Santiago
APPLICANT: Barredo, Jose L.
APPLICANT: Von Doehren, Hans
APPLICANT: Palissa, Harriet
APPLICANT: Van Liempt, Henk
APPLICANT: Montenegro, Eduardo P.
TITLE OF INVENTION: A Method for Influencing Beta-Lactam
TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
TITLE OF INVENTION: Quantities of ACV Synthetase
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: Illinois

COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,617A
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION: 97,157
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 11601 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Acremonium chrysogenum
FEATURE:
NAME/KEY: CDS
LOCATION: 388..11526
OTHER INFORMATION: /function= "Enzyme"
OTHER INFORMATION: /product= "Acv Synthetase"
FEATURE:
NAME/KEY: CDS
LOCATION: 8050..8052
OTHER INFORMATION:
US-08-222-617A-24 /note= "NNN=AGU, AGC, UCU, UCC, UCA, or UCG"

Query Match 53.5%; Score 18.2; DB 2; Length 11601;
Best Local Similarity 87.0%; Pred. No. 19;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 10 ACTTATGATCGTACTGTGAGA 32
Db 11163 ACTTATGATCGTACTGTGACA 11185

RESULT 7
US-09-280-116-189
Sequence 189, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035800/176965
CURRENT APPLICATION NUMBER: US/09/280,116A
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 189
LENGTH: 507
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: astacin/m 12a metalloproteases
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(507)
OTHER INFORMATION: n = a, t, c or g
US-09-280-116-189

Query Match 51.8%; Score 17.6; DB 4; Length 507;
Best Local Similarity 83.3%; Pred. No. 18;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 CTCTACTTATGATCGGTACTGTG 29
Db 11163 CTCTACTTATGATCGGTACTGTG 29

Db 44 CTCTCTTATAGATCGTACTGTG 67
RESULT 8
US-09-280-116-191
Sequence 191, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035800/176965
CURRENT APPLICATION NUMBER: US/09/280,116A
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 191
LENGTH: 933
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: astacin/m 12a metalloproteases
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(933)
OTHER INFORMATION: n = a, t, c or g
US-09-280-116-191

Query Match 51.8%; Score 17.6; DB 4; Length 933;
Best Local Similarity 83.3%; Pred. No. 20;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 CTCTACTTATGATCGGTACTGTG 29
Db 289 CTCTCTTATAGATCGTACTGTG 312

RESULT 9
US-08-974-022-1
Sequence 1, Application US/08974022
Patent No. 6015938
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,022
FILING DATE: 12-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326
US-08-974-022-1

Query Match 51.2%; Score 17.4; DB 3; Length 2432;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TCCTCTACTATGATCGTACTGTGA 30
DB 1602 TCCTCTCTTATGATGATGACTCAGA 1628

RESULT 10
US-08-795-445A-1
Sequence 1, Application US/08795445A
Patent No. 6284485
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,445A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326
US-08-795-445A-1

Query Match 51.2%; Score 17.4; DB 4; Length 2432;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TCCTCTACTATGATCGTACTGTGA 30
DB 1602 TCCTCTCTTATGATGATGACTCAGA 1628

RESULT 11
US-08-795-447A-1

Sequence 1, Application US/08795447A
Patent No. 6284728
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: Osteoprotegerin
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Amgen Inc.
STREET: One Amgen Center Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91362-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,447A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378D2
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 124..1326
US-08-795-447A-1

Query Match 51.2%; Score 17.4; DB 4; Length 2432;
Best Local Similarity 77.8%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 TCCTCTACTATGATCGTACTGTGA 30
DB 1602 TCCTCTCTTATGATGATGACTCAGA 1628

RESULT 12
US-08-974-186-1
Sequence 1, Application US/08974186
Patent No. 6284740
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

1 APPLICATION NUMBER: US/08/974,186
2
3 FILING DATE:
4 CLASSIFICATION:
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: 08/577,788
7
8 FILING DATE:
9
10 ATTORNEY/AGENT INFORMATION:
11 NAME: WINTER, ROBERT B.
12 REFERENCE/DOCKET NUMBER: A-378
13
14 INFORMATION FOR SEQ ID NO: 1:
15
16 SEQUENCE CHARACTERISTICS:
17
18 LENGTH: 2432 base pairs
19 TYPE: nucleic acid
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22
23 MOLECULE TYPE: CDNA
24
25 FEATURE:
26
27 NAME/KEY: CDS
28 LOCATION: 124..1326
29
30 US-08-974-186-1

```

Query Match	51.2%	Score 17.4;	DB 4;	Length 2432;
Best Local Similarity	77.8%;	Pred. No. 32;		
Matches 21; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

```

QY      4  TCCTCTACTTATCGATCGGTACTGTGA  30
          ||||| ||||| ||||| |||||
Db      1602 TCCTCTCCTTATGATGAGTACTCAGA  1628

```

RESULT 13
US-08-795-446B-1
; Sequence 1, Application US/08795446B

GENERAL INFORMATION: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:

```

? ZIP: 91320-1789
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: ParentIn Release #1.0, Version #1.3D
? COMPANY: Advantec International, Inc.

```

US-08-795-446B-1

Query Match	51.2%	Score 17.4;	DB 4;	Length 2432;
Best Local Similarity	77.8%	Pred. No. 32;		
Matches 21; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

QY 4 TCCTCTACTTATCGATCGGTA CTGTGA 30
||||| ||| ||||| ||
Db 1602 TCCTCTCTTATGATGAGTACTCAGA 1628

```

RESULT 14
US-08-706-945D-123
Sequence 123. Application US/08706945D
Patent No. 6369027
GENERAL INFORMATION:
APPLICANT: Boyle, William
APPLICANT: Lacey, David
APPLICANT: Calzone, Frank
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: Osteoprotegerin
FILE REFERENCE: A-37831P
CURRENT APPLICATION NUMBER: US/08/706,945D
CURRENT FILING DATE: 1996-09-03
PRIOR APPLICATION NUMBER: 08/577,788
PRIOR FILING DATE: 1995-12-22
NUMBER OF SEQ ID NOS: 145
SOFTWARE: PatentIn version 3.1
SEQ ID NO 123
LENGTH: 2432
TYPE: DNA
ORGANISM: Rattus rattus
FEATURE:
NAME/KEY: CDS
LOCATION: (134)..(1326)
OTHER INFORMATION:
US-08-706-945D-123

```

Query Match	51.2%	Score 17.4;	DB 4;	Length 2432;
Best Local Similarity	77.8%;	Pred. No. 32;		
Matches 21; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

QY 4 TCCTCTACTTATCGATCGGTACTGTGA 30
 |||||
Db 1602 TCCTCTCCTTAATGATGAGTACTCAGA 1628

US-08-364-081-2/c
Sequence 2, Application US/08364081
Patent No. 5585464
GENERAL INFORMATION:
APPLICANT: Ramesh K. Prakash
TITLE OF INVENTION: Recombinant Antigen for Diagnosing
TITLE OF INVENTION: Rheumatoid Arthritis
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thorpe, No. 5585464th & Western
STREET: 9035 South 700 East, Suite 200
City: Sandy
STATE: Utah
COUNTRY: USA
ZIP: 84070
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 720 kb storage
COMPUTER: AST Advantage NB-SX20
OPERATING SYSTEM: DOS 6.2
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/364,081
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/019,780
FILING DATE: 19-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Alan J. Howarth
REGISTRATION NUMBER: 36,553
REFERENCE/DOCKET NUMBER: T781CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (801)566-6633
TELEFAX: (801)566-0750
INFORMATION FOR SEO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 993 nucleotides
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-364-081-2

Query Match 50.6%; Score 17.2; DB 1; Length 993;
Best Local Similarity 86.4%; Pred. No. 32;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 10 ACTTATCGATCGTACTGTGAG 31
Db 265 ACTTCTCGATTGGAACGTGTGAG 244

Search completed: May 11, 2003, 03:07:47
Job time : 18.6341 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:08:16 / Search time 33.6591 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-15

Perfect score: 34

Sequence: 1 gactcctctactatcgatcgatcgatcgatcgagaca 34

Scoring table: IDENTITY_NUC

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PTC_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PTCUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	34	9 US-09-963-803-15	Sequence 15, Appl
2	19.2	56.5	543	9 US-09-738-626-619	Sequence 619, App
3	19.2	56.5	612	9 US-09-738-626-620	Sequence 620, App
4	19.2	56.5	3309400	9 US-09-738-626-61	Sequence 1, Appl
5	18.6	54.7	750	12 US-10-039-785-62	Sequence 62, Appl
6	18.2	53.5	403	10 US-09-783-590-294	Sequence 294, App
7	18.2	53.5	536165	9 US-09-939-964-1	Sequence 1, Appl
8	18	52.9	4259	8 US-08-927-939-30	Sequence 30, Appl
9	18	52.9	1503841	9 US-09-945-807-1	Sequence 1, Appl
10	18	52.9	1503841	10 US-09-795-668-1	Sequence 1, Appl
11	18	52.9	1503841	10 US-09-795-668-1	Sequence 1, Appl
12	17.8	52.4	371	9 US-09-918-995-3042	Sequence 3042, A
13	17.8	52.4	3371	9 US-10-028-072-35	Sequence 35, Appl
14	17.8	52.4	3371	9 US-10-121-049-35	Sequence 35, Appl
15	17.8	52.4	3371	9 US-10-123-904-35	Sequence 35, Appl
16	17.8	52.4	3371	9 US-10-140-470-35	Sequence 35, Appl
17	17.8	52.4	3371	9 US-10-175-746-35	Sequence 35, Appl
18	17.8	52.4	3371	9 US-10-176-918-35	Sequence 35, Appl
19	17.8	52.4	3371	9 US-10-176-921-35	Sequence 35, Appl

C 20	17.8	52.4	3371	9	US-10-137-865-35	Sequence 35, Appl
C 21	17.8	52.4	3371	9	US-10-140-474-35	Sequence 35, Appl
C 22	17.8	52.4	3371	9	US-10-142-431-35	Sequence 35, Appl
C 23	17.8	52.4	3371	9	US-10-143-114-35	Sequence 35, Appl
C 24	17.8	52.4	3371	9	US-10-140-002-35	Sequence 35, Appl
C 25	17.8	52.4	3371	9	US-10-142-419-35	Sequence 35, Appl
C 26	17.8	52.4	3371	9	US-10-123-262-35	Sequence 35, Appl
C 27	17.8	52.4	3371	9	US-10-142-423-35	Sequence 35, Appl
C 28	17.8	52.4	3371	9	US-10-121-050-35	Sequence 35, Appl
C 29	17.8	52.4	3371	9	US-10-141-755-35	Sequence 35, Appl
C 30	17.8	52.4	3371	9	US-10-143-032-35	Sequence 35, Appl
C 31	17.8	52.4	3371	9	US-10-123-108-35	Sequence 35, Appl
C 32	17.8	52.4	3371	9	US-10-123-236-35	Sequence 35, Appl
C 33	17.8	52.4	3371	9	US-10-123-261-35	Sequence 35, Appl
C 34	17.8	52.4	3371	9	US-10-140-928-35	Sequence 35, Appl
C 35	17.8	52.4	3371	9	US-10-140-928-35	Sequence 35, Appl
C 36	17.8	52.4	3371	9	US-10-121-045-35	Sequence 35, Appl
C 37	17.8	52.4	3371	9	US-10-123-903-35	Sequence 35, Appl
C 38	17.8	52.4	3371	9	US-10-124-819-35	Sequence 35, Appl
C 39	17.8	52.4	3371	9	US-10-140-925-35	Sequence 35, Appl
C 40	17.8	52.4	3371	9	US-10-160-498-35	Sequence 35, Appl
C 41	17.8	52.4	3371	9	US-10-121-041-35	Sequence 35, Appl
C 42	17.8	52.4	3371	9	US-10-121-043-35	Sequence 35, Appl
C 43	17.8	52.4	3371	9	US-10-121-043-35	Sequence 35, Appl
C 44	17.8	52.4	3371	9	US-10-121-047-35	Sequence 35, Appl
C 45	17.8	52.4	3371	9	US-10-121-047-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1
US-09-963-803-15
Sequence 15, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15
LENGTH: 34
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Guide desoxynucleotide building
US-09-963-803-15
Query Match 100.0%; Score 34; DB 9; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 GACTCCTCTACTATCGATCGATCGATCGATCGAGACA 34
1 GACTCCTCTACTATCGATCGATCGATCGATCGAGACA 34
RESULT 2
US-09-738-626-619
Sequence 619, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI

```

APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 619
LENGTH: 543
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-619

```

```

Query Match 56.5%; Score 19.2; DB 9; Length 543;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 1 GACTCTCTACTTATGATCGTACTGTGAGA 32
DB 441 GACTGCTCTCTGATCGATCAGTACGTGAGA 472

RESULT 3
US-09-738-626-620/c
Sequence 620, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 620
LENGTH: 612
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-620

```

```

Query Match 56.5%; Score 19.2; DB 9; Length 612;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 GACTCTCTACTTATGATCGTACTGTGAGA 32

```

```

DB 154 GACTGCTCTCTGATCGATCAGTACGTGAGA 123

```

```

RESULT 4
US-09-738-626-1
Sequence 62, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

```

```

Query Match 56.5%; Score 19.2; DB 9; Length 3309400;
Best Local Similarity 75.0%; Pred. No. 1,1e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 1 GACTCTCTACTTATGATCGTACTGTGAGA 32
DB 571214 GACTGCTCTCTGATCGATCAGTACGTGAGA 571245

```

```

RESULT 5
US-10-039-785-62/c
Sequence 62, Application US/10039785
Patent No. US20020067646A1
GENERAL INFORMATION:
APPLICANT: Salcedo et al.
TITLE OF INVENTION: Antibodies that Immunospecifically Bind to TRAIL
FILE REFERENCE: PFS50
CURRENT APPLICATION NUMBER: US/10/039,785
CURRENT FILING DATE: 2002-05-07
PRIOR APPLICATION NUMBER: 60/369,860
PRIOR FILING DATE: 2002-04-05
PRIOR APPLICATION NUMBER: 60/341,237
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: 60/331,310
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/331,044
PRIOR FILING DATE: 2001-11-07
PRIOR APPLICATION NUMBER: 60/327,364
PRIOR FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 60/323,807
PRIOR FILING DATE: 2001-09-21
PRIOR APPLICATION NUMBER: 60/309,176
PRIOR FILING DATE: 2001-08-02
PRIOR APPLICATION NUMBER: 60/294,981
PRIOR FILING DATE: 2001-06-04

```

PRIOR APPLICATION NUMBER: 60/293,473
PRIOR FILING DATE: 2001-05-25
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 62
LENGTH: 750
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: DNA encoding T1015A02 scfv
US-10-039-785-62

Query Match 54.7%; Score 18.6; DB 12; Length 750;
Best Local Similarity 72.7%; Pred. No. 43;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGTACTGTGAG 33
DB 235 GATTCTGAGAGTATCCATCGATACGTAAGTC 203

RESULT 6

US-09-783-590-294
Sequence 294, Application US/09783590
Patent No. US20020110850A1
GENERAL INFORMATION:
APPLICANT: Dillon, Patrick J.
APPLICANT: Haseltine, William A.
APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
FILE REFERENCE: PO-16.2C1
CURRENT APPLICATION NUMBER: US/09/783,590
CURRENT FILING DATE: 2000-02-15
PRIOR APPLICATION NUMBER: 08/420,856
PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 294
LENGTH: 403
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (10)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (141)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (327)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (400)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-294

Query Match 53.5%; Score 18.2; DB 10; Length 403;
Best Local Similarity 74.2%; Pred. No. 59;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGTACTGTGAG 31
DB 337 GAATTCGAACTTTTCAAGGGGTACTGTGAG 367

RESULT 7

US-09-939-964-1/C
Sequence 1, Application US/09939964
Publication No. US20030054522A1

GENERAL INFORMATION:
APPLICANT: Rosenthal, Andre
APPLICANT: Freiberg, Christoph
APPLICANT: Perret, Xavier Philippe
APPLICANT: Broughton, William John
TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
FILE REFERENCE: CARP0068
CURRENT APPLICATION NUMBER: US/09/939,964
CURRENT FILING DATE: 2001-08-27
PRIOR APPLICATION NUMBER: 09/214,808
PRIOR FILING DATE: 1999-06-22
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 536165
TYPE: DNA
ORGANISM: Rhizobium
US-09-939-964-1

Query Match 53.5%; Score 18.2; DB 9; Length 536165;
Best Local Similarity 74.2%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGTACTGTGAG 31
DB 14804 GACTCCTCTGTTTTCGATCGGCGGCTTAG 14774

RESULT 8

US-08-927-939-30/C
Sequence 30, Application US/08927939
Patent No. US20010006640A1
GENERAL INFORMATION:
APPLICANT: Grainger, David J.
APPLICANT: Tatalick, Lauren Marie
TITLE OF INVENTION: Compounds and methods to inhibit or
FILE REFERENCE: 295.022US1
CURRENT APPLICATION NUMBER: US/08/927,939
CURRENT FILING DATE: 1997-09-11
NUMBER OF SEQ ID NOS: 83
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 30
LENGTH: 4259
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (2070)...(2130)
NAME/KEY: CDS
LOCATION: (2669)...(2795)
NAME/KEY: CDS
LOCATION: (2990)...(3079)
NAME/KEY: CDS
LOCATION: (3491)...(3506)
US-08-927-939-30

Query Match 52.9%; Score 18; DB 8; Length 4259;
Best Local Similarity 70.6%; Pred. No. 1.2e+02;
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 GACTCCTCTACTTATCGATCGTACTGTGAG 34
DB 1165 GAACCTCTACCCACCAATGGAATGGGAACA 1132

RESULT 9

US-09-946-807-1/C
Sequence 1, Application US/09946807
Patent No. US20020165144A1
GENERAL INFORMATION:
APPLICANT: Stefansson, Hreinn

Query	1	GACTCCTCTACTTATCGATGGCTACTGTGAGACA	34
Db	183456	GAATTCCTCTATTATAAATGTGCTACTGGAGAAA	183423
Best Local Similarity 52.9%; Score 18; DB 9; Length 1503841;			
Matches 24; Conservative 0; Mismatches 10; Indels 0; Gaps 0;			

5

	Query Match	52.9%	Score 18:	DB 10;	Length 1503841;
	Blast Local Similarity	70.6%;	Pred. No.	3.1e+07;	
	Matches 24;	Conservative	0;	Mismatches 10;	Indels 0;
	Gaps				
OY	1 GACTCTCTACTTATCGATCGGAGACTGTGAACA	34			
bD	183456 GAATTCCCTATTTTAATAATTGGTACTGGGAGAAAA	183423			

```

      i ORGANISM: Homo sapiens
US-09-918-995-30042

Query Match          52.4%; Score 17.8; DB 9; Length 371;
Best Local Similarity 75.9%; Pred.No. 89;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0.

OY      5 CCTCTACTATGCGTGGTACTGTGCAGAC 33
       |||||
Db      191 CCTTACTTATAGATCTGCACGGTTGGGC 219

RESULT 13
US-10-028-072-35/C
; Sequence 35, Application US/10028072
; Publication No. US2003004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey J.
; APPLICANT: Gudowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/028,072
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059836
PRIOR FILING DATE: 1997-09-24
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062285
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062814
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/062816
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063045
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063082

```

PRIOR FILING DATE: 1997-10-31
 PRIOR APPLICATION NUMBER: 60/063127
 PRIOR FILING DATE: 1997-10-24
 PRIOR APPLICATION NUMBER: 60/063327
 PRIOR FILING DATE: 1997-10-27
 PRIOR APPLICATION NUMBER: 60/063329
 PRIOR FILING DATE: 1997-10-27
 PRIOR APPLICATION NUMBER: 60/063550
 PRIOR FILING DATE: 1997-10-28
 PRIOR APPLICATION NUMBER: 60/063561
 PRIOR FILING DATE: 1997-10-28
 PRIOR APPLICATION NUMBER: 60/063704
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063733
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063735
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063738
 PRIOR FILING DATE: 1997-10-29
 PRIOR APPLICATION NUMBER: 60/063755
 PRIOR FILING DATE: 1997-10-17
 PRIOR APPLICATION NUMBER: 60/064248
 PRIOR FILING DATE: 1997-11-03
 PRIOR APPLICATION NUMBER: 60/064809
 PRIOR FILING DATE: 1997-11-07
 PRIOR APPLICATION NUMBER: 60/065186
 PRIOR FILING DATE: 1997-11-12
 PRIOR APPLICATION NUMBER: 60/065846
 PRIOR FILING DATE: 1997-11-17
 PRIOR APPLICATION NUMBER: 60/066364
 PRIOR FILING DATE: 1997-11-21
 PRIOR APPLICATION NUMBER: 60/066453
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/066511
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/066770
 PRIOR FILING DATE: 1997-11-24
 PRIOR APPLICATION NUMBER: 60/069212
 PRIOR FILING DATE: 1997-12-11
 PRIOR APPLICATION NUMBER: 60/069278
 PRIOR FILING DATE: 1997-12-11
 PRIOR APPLICATION NUMBER: 60/069334
 PRIOR FILING DATE: 1997-12-11
 PRIOR APPLICATION NUMBER: 60/069694
 PRIOR FILING DATE: 1997-12-16
 PRIOR APPLICATION NUMBER: 60/072320
 PRIOR FILING DATE: 1998-01-23
 PRIOR APPLICATION NUMBER: 60/073612
 PRIOR FILING DATE: 1998-02-04
 PRIOR APPLICATION NUMBER: 60/074086
 PRIOR FILING DATE: 1998-02-09
 PRIOR APPLICATION NUMBER: 60/074092
 PRIOR FILING DATE: 1998-02-09
 PRIOR APPLICATION NUMBER: 60/077791
 PRIOR FILING DATE: 1998-03-12
 PRIOR APPLICATION NUMBER: 60/078910
 PRIOR FILING DATE: 1998-03-20
 PRIOR APPLICATION NUMBER: 60/079294
 PRIOR FILING DATE: 1998-03-25
 PRIOR APPLICATION NUMBER: 60/079663
 PRIOR FILING DATE: 1998-02-27
 PRIOR APPLICATION NUMBER: 60/079728
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: 60/080165
 PRIOR FILING DATE: 1998-03-31
 PRIOR APPLICATION NUMBER: 60/081203
 PRIOR FILING DATE: 1998-04-09
 PRIOR APPLICATION NUMBER: 60/081229
 PRIOR FILING DATE: 1998-04-09
 PRIOR APPLICATION NUMBER: 60/081695
 PRIOR FILING DATE: 1998-04-14
 PRIOR APPLICATION NUMBER: 60/081817
 PRIOR FILING DATE: 1998-04-15

PRIOR APPLICATION NUMBER: 60/081818
 PRIOR FILING DATE: 1998-04-15
 PRIOR APPLICATION NUMBER: 60/082999
 PRIOR FILING DATE: 1998-04-24
 PRIOR APPLICATION NUMBER: 60/083322
 PRIOR FILING DATE: 1998-04-28
 PRIOR APPLICATION NUMBER: 60/083545
 PRIOR FILING DATE: 1998-04-29
 PRIOR APPLICATION NUMBER: 60/084600
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084627
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/084637
 PRIOR FILING DATE: 1998-05-07
 PRIOR APPLICATION NUMBER: 60/085149
 PRIOR FILING DATE: 1998-05-12
 PRIOR APPLICATION NUMBER: 60/085323
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085338
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085339
 PRIOR FILING DATE: 1998-05-13
 PRIOR APPLICATION NUMBER: 60/085579
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085697
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/085704
 PRIOR FILING DATE: 1998-05-15
 PRIOR APPLICATION NUMBER: 60/086414
 PRIOR FILING DATE: 1998-05-22
 PRIOR APPLICATION NUMBER: 60/086430
 PRIOR FILING DATE: 1998-05-22
 PRIOR APPLICATION NUMBER: 60/087106
 PRIOR FILING DATE: 1998-05-28
 PRIOR APPLICATION NUMBER: 60/088026
 PRIOR FILING DATE: 1998-06-04
 PRIOR APPLICATION NUMBER: 60/088730
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088741
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088810
 PRIOR FILING DATE: 1998-06-10
 PRIOR APPLICATION NUMBER: 60/088858
 PRIOR FILING DATE: 1998-06-11
 PRIOR APPLICATION NUMBER: 60/089532
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089599
 PRIOR FILING DATE: 1998-06-17
 PRIOR APPLICATION NUMBER: 60/089907
 PRIOR FILING DATE: 1998-06-18
 PRIOR APPLICATION NUMBER: 60/089947
 PRIOR FILING DATE: 1998-06-19
 PRIOR APPLICATION NUMBER: 60/090349
 PRIOR FILING DATE: 1998-06-23
 PRIOR APPLICATION NUMBER: 60/090429
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090445
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090538
 PRIOR FILING DATE: 1998-06-24
 PRIOR APPLICATION NUMBER: 60/090863
 PRIOR FILING DATE: 1998-06-26
 PRIOR APPLICATION NUMBER: 60/091360
 PRIOR FILING DATE: 1998-07-01
 PRIOR APPLICATION NUMBER: 60/091519
 PRIOR FILING DATE: 1998-07-02
 PRIOR APPLICATION NUMBER: 60/091982
 PRIOR FILING DATE: 1998-07-07

Query Match 52.4%; Score 17.8; DB 9; Length 3371;
 Best Local Similarity 75.9%; Pred. No. 1.4e+02;
 Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 6 CTCTACTTATCGATCGGTACTGTGAGACA 34
Db 3291 CTCTATCTACCCACTCTGTACTTAGAGACA 3263

RESULT 14

US-10-121-049-35/C
Sequence 35, Application US/10121049
Publication No. US2003002239A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C17
CURRENT FILING DATE: 2002-04-12
CURRENT APPLICATION NUMBER: US/10/121,049
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 35
LENGTH: 3371
TYPE: DNA
ORGANISM: Homo Sapien
US-10-121-049-35

Query Match 52.4%; Score 17.8; DB 9; Length 3371;
Best Local Similarity 75.9%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 6 CTCTACTTATCGATCGGTACTGTGAGACA 34
Db 3291 CTCTATCTACCCACTCTGTACTTAGAGACA 3263

RESULT 15

US-10-123-904-35/C
Sequence 35, Application US/10123904
Publication No. US20030022328A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C54
CURRENT FILING DATE: 2002-04-16
CURRENT APPLICATION NUMBER: US/10/123,904
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 35
LENGTH: 3371
TYPE: DNA
ORGANISM: Homo Sapien
US-10-123-904-35

FILE REFERENCE: P3330R1C54
CURRENT APPLICATION NUMBER: US/10/123,904
CURRENT FILING DATE: 2002-04-16
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 35
LENGTH: 3371
TYPE: DNA
ORGANISM: Homo Sapien
US-10-123-904-35

Query Match 52.4%; Score 17.8; DB 9; Length 3371;
Best Local Similarity 75.9%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 6 CTCTACTTATCGATCGGTACTGTGAGACA 34
Db 3291 CTCTATCTACCCACTCTGTACTTAGAGACA 3263

Search completed: May 11, 2003, 06:27:57
Job time : 1007.66 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 546.556 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-15

Sequence: 1 gactccttacttatcgatcgttactgtgagaca 34

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```

1:  EST:.*
2:  em_estha:.*
3:  em_esthum:.*
4:  em_estlin:.*
5:  em_estnum:.*
6:  em_estov:.*
7:  em_estpy:.*
8:  em_estro:.*
9:  em_estt:.*
10: gb_est1:.*
11: gb_est2:.*
12: gb_estc:.*
13: gb_estd:.*
14: gb_este:.*
15: em_estum:.*
16: em_estcom:.*
17: gb_g88:.*
18: em_g88_hum:.*
19: em_g88_inv:.*
20: em_g88_pln:.*
21: em_g88_vtr:.*
22: em_g88_fun:.*
23: em_g88_mam:.*
24: em_g88_mus:.*
25: em_g88_other:.*
26: em_g88_pio:.*
27: em_g88_rtd:.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21.2	62.4	665	17	A2607860	A2607800 IM0430M22
2	20.8	61.0	589	13	BMO96067	BMO96067 F430C09.3
3	20.4	60.0	599	13	B1672108	B1672108 F434E03.1
4	20.4	60.0	660	14	BQ109731	BQ109731 Image8C
5	20.2	59.4	702	13	B1823138	B1823138 6010035F05
6	20.2	58.8	136	12	BG833308	BG833308 951004F05

C	7	19.8	58.2	349	12	BF542321	BF542321	UT-R-C2P
C	8	19.8	58.2	451	9	AL387696	AL387696	MEBC44B09
C	9	19.8	58.2	502	13	BM209528	BM209528	C0649B07-
C	10	19.8	58.2	514	14	BQ559633	BQ559633	H4059D12-
C	11	19.8	58.2	524	10	BM764031	BM764031	BF8764031
C	12	19.8	58.2	530	10	BB752196	BB752196	BB752196
C	13	19.8	58.2	536	17	AZ331947	AZ331947	BM751947
C	14	19.8	58.2	611	17	AZ831669	AZ831669	1M0060J06
C	15	19.8	58.2	615	14	BM4A5593	BM4A5593	2M0111H02
C	16	19.6	57.6	186	14	BM816979	BM816979	UT-M-EMO-
C	17	19.6	57.6	373	13	BF063890	BF063890	KC2-BN024
C	18	19.6	57.6	771	12	BF104508	BF104508	H0044190
C	19	19.6	57.6	979	10	BE293067	BE293067	601647475
C	20	19.6	57.6	1002	12	BB875770	BB875770	601106072
C	21	19.4	57.1	224	17	AZ007407	AZ007407	601487228
C	22	19.4	57.1	229	10	BB859348	BB859348	RPCI-23-3
C	23	19.4	57.1	665	14	BQ086932	BQ086932	BB593348
C	24	19.4	57.1	690	17	BM265000	BM265000	BB593348
C	25	19.4	57.1	713	13	BI521708	BI521708	Cr1-10 CO
C	26	19.2	56.5	213	10	AM445625	AM445625	BH255000
C	27	19.2	56.5	249	9	A1616637	A1616637	CH230-66E
C	28	19.2	56.5	304	17	CNS00DP6	AL0844576	BI571708
C	29	19.2	56.5	332	13	BM573163	BM573163	603081645
C	30	19.2	56.5	332	13	BM573789	BM573789	81979 MAR
C	31	19.2	56.5	374	17	BH348403	BH348403	zenn0382.
C	32	19.2	56.5	380	13	BM157390	BM157390	AL616637
C	33	19.2	56.5	389	13	BM342056	BM342056	AL0844576
C	34	19.2	56.5	414	14	BQ073668	BQ073668	ArabiDp5
C	35	19.2	56.5	445	13	BM316704	BM316704	BM573163
C	36	19.2	56.5	484	14	D32654	D32654	fK99C02.Y
C	37	19.2	56.5	501	17	AO148607	AO148607	EY03H03.Y
C	38	19.2	56.5	505	10	AM777682	AM777682	BH348403
C	39	19.2	56.5	524	9	AA593744	AA593744	CH230-184
C	40	19.2	56.5	543	17	AZ289841	AZ289841	BM157390
C	41	19.2	56.5	547	14	BQ133481	BQ133481	fK42d06.Y
C	42	19.2	56.5	572	17	AZ340572	AZ340572	BM542056
C	43	19.2	56.5	587	17	AO673369	AO673369	fW68d06.X
C	44	19.2	56.5	597	10	AM777484	AM777484	fK33e01.X
C	45	19.2	56.5	611	14	BM777858	BM777858	fW33g01.s

ALIGNMENTS

RESULT 1	
AZ607800	
LOCUS	665 bp DNA linear GSS 13-DEC-2000
DEFINITION	1M0430M22F Mouse 10kb plasmid UGCG1M library Mus musculus genomic clone UGCG1M0430M22 F, DNA sequence.
ACCESSION	AZ607800
VERSION	AZ607800.1
KEYWORDS	GI:11729990
SOURCE	GSS.
ORGANISM	house mouse.
REFERENCE	Mus musculus
AUTHORS	Emmariyoti; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurgnachti; Muridae; Murinae; Mus. 1 (bases 1 to 665)
TITLE	Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A. and Wright, D., Weiss, R.
JOURNAL COMMENT	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts Unpublished (2000) Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: ddum@genetics.utah.edu Insert Length: 10000 Std Error: 0.00

contact Zhiyuan Gong for further information on this library
(National University of Singapore, Department of Biological
Sciences, Lower Kent Ridge Road, Singapore 119260).
Seq primer: T7 from Gibco
High quality sequence stop: 463.

FEATURES

Source

Location/Qualifiers

1. 599

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone="5152637"

/clone_1lb="Gong zebrafish testis"

/sex="male"

/dev_stage="4-5 month"

/lab_host="DH10B (phage-resistant)"

/note="Organ: testis (pooled); Vector: pBluescript SK-;

Site 1: XhoI; Site 2: BcoRI; Poly A+ RNA was isolated from

the testes of 31 male adult zebrafish (4-5 month old).

CDNAs were made using oligo-dT primers and inserted into

lambda Zap II vector (Stratagene) by Dr. Z. Gong, in vivo

mass-excised to pBluescript SK- following the Washington

University protocol

(http://genome.wustl.edu/est/lambda_protocol.shtml).

Please contact Zhiyuan Gong for further information on

this library (National University of Singapore,

Department of Biological Sciences, Lower Kent Ridge Road,

Singapore 119260)."

BASE COUNT

198 a 82 c 117 g 202 t

ORIGIN

Query Match 60.0%; Score 20.4; DB 13; Length 599;
Best Local Similarity 80.0%; Pred. No. 4.6e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 CTCTCTACTTATCGATCGGTAAGTGTGAGA 32
DB 386 CTCTCTCTCTTTTGTATCGAAGTGTGAGA 357

RESULT 4

BO109731 660 bp mRNA linear EST 16-APR-2002

LOCUS

imagec_7_2001/bnm420bdf41.x1 Gong zebrafish testis Danio rerio

DEFINITION

CDNA clone IMAGE:5152637 3', mRNA sequence.

ACCESSION

BO109731

VERSION

BO109731.1 GI:20159385

KEYWORDS

EST.

SOURCE

zebrafish.

ORGANISM

Danio rerio

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.

AUTHORS

Kale, P. I., Harshb, T. J., Folte, P. A., Nelson, D. O., Sanders, C. G. and
Prange, C. K.

TITLE

The I.M.A.G.E. Consortium quality control effort: clone

JOURNAL

Unpublished (2001)

COMMENT

Contact: Prange CK

FEATURES

The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.llnl.gov
This read has been produced as part of the I.M.A.G.E. Consortium
quality control effort. High quality sequence is defined as having
100 or more base pairs with a phased quality value of 20 or greater,
where a sliding window of 4 base pairs with a phased quality value
of 15 or greater marks the beginning and end of the sequence. For
information on obtaining this clone, please contact
info@image.llnl.gov.
Plate: LHAM11377 row: k column: 6
Seq primer: -21m13
High quality sequence stop: 660.
Location/Qualifiers

Source

1. 660

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone="IMAGE:5152637"

/clone_1lb="Gong zebrafish testis"

/sex="male"

/dev_stage="4-5 month"

/lab_host="DH10B (phage-resistant)"

/note="Organ: testis (pooled); Vector: pBluescript SK-;

Site 1: XhoI; Site 2: BcoRI; Poly A+ RNA was isolated from

the testes of 31 male adult zebrafish (4-5 month old).

CDNAs were made using oligo-dT primers and inserted into

lambda Zap II vector (Stratagene) by Dr. Z. Gong, in vivo

mass-excised to pBluescript SK- following the Washington

University protocol

(http://genome.wustl.edu/est/lambda_protocol.shtml).

Please contact Zhiyuan Gong for further information on

this library (National University of Singapore,

Department of Biological Sciences, Lower Kent Ridge Road,

Singapore 119260)."

BASE COUNT

217 a 88 c 127 g 228 t

ORIGIN

Query Match 60.0%; Score 20.4; DB 14; Length 660;
Best Local Similarity 80.0%; Pred. No. 4.8e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 3 CTCTCTACTTATCGATCGGTAAGTGTGAGA 32
DB 413 CTCTCTCTCTTTTGTATCGAAGTGTGAGA 384

RESULT 5

BI823138 702 bp mRNA linear EST 04-OCT-2001

LOCUS

603039579F1 NIH_MGC_115 Homo sapiens CDNA clone IMAGE:5180730 5',

DEFINITION

mRNA sequence.

ACCESSION

BI823138

VERSION

BI823138.1 GI:15934688

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

TITLE

Unpublished (1999)

JOURNAL

Contact: Robert Strausberg, Ph.D.

COMMENT

Email: csabbs-remail.nih.gov

FEATURES

Tissue Procurement: Life Technologies, Inc.

Source

CDNA Library Preparation: Life Technologies, Inc.

Source

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

Source

DNA Sequencing by: Incyte Genomics, Inc.

Source

Clone distribution: MGC clone distribution information can be

Source

found through the I.M.A.G.E. Consortium/LNL at:

Source

http://image.llnl.gov

Source

Plate: LHAM11450 row: m column: 19

Source

High quality sequence stop: 702.

Source

Location/Qualifiers

Source

1. 702

Source

/organism="Homo sapiens"

Source

/db_xref="taxon:9606"

Source

/clone="IMAGE:5180730"

Source

/clone_1lb="NIH_MGC_115"

Source

/lab_host="DH10B"

Source

/note="Organ: pooled brain, lung, testis; Vector:

Source

pcwv-sPORT6; Site 1: NotI; Site 2: BcoRI (destroyed); RNA

Source

source anonymous pool of 6 male brains, age range 23-27; 1

Source

male lung, age 27; and 1 male testis, age 69. Library is

Source

oligo-dT primed and directionally cloned (BcoRI site is

Source

destroyed upon cloning). Average insert size 1.8 kb,

Source

insert size range 1-3 kb. Library is normalized and

enriched for full-length clones and was constructed by C. Gruber (Invitrogen) .. Research Genetics tracking code 021. Note: this is a NIH MGC Library."

BASE COUNT

209 a 155 c 160 g 178 t

Query Match 59.4%; Score 20.2; DB 13; Length 702;
Best Local Similarity 75.8%; Pred. No. 5.9e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 2 ACTGCTTACTTATGATCGTACTGTGACA 34
670 ACTGCTGTTATCAATGTTATATATGACA 702

RESULT 6

LOCUS

951004F05.X2 951 - BMS tissue from Walbot Lab (GR) Zea mays cDNA,
mRNA sequence.

ACCESSION BG833308.1 GI:14191078

KEYWORDS

EST.

SOURCE

Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 8221
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 951004 row: F column: 05.
Location/Qualifiers

FEATURES

source

1. 136
/organism="Zea mays"
/cultivar="BMS (Black Mexican Sweet)"
/db_xref="taxon:4577"
/clone_id="951 - BMS tissue from Walbot Lab (GR)"
/tissue_type="suspension culture"
/dev_stage="mixed logarithmic and stationary growth
phases"
/lab_host="DH10B"
/note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI. The
library was prepared by George Rudenko using poly (A)
selected RNA and Universal Riboclone cDNA Synthesis System
(Promega). cDNA was synthesized using both random and
oligo(dT) primers in separate reactions and equipped with
EcoRI adaptors. Library was size-fractionated on agarose
gels (for insert size >400bp) and non-directionally cloned
into EcoRI-digested pUC19 vector. Blue/white selection on
carbenicillin-containing plates was used to recover
positive clones."

BASE COUNT

34 a 35 c 31 g 36 t

Query Match

Best Local Similarity 58.8%; Score 20; DB 12; Length 136;
Pred. No. 4.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 5 CCTCTACTTATGATCGTACTGTGACA 32

76 CCTCTACTTATGATCGTACTGTGACA 103

RESULT 7
BF542321/c 349 bp. mRNA linear EST 11-DEC-2000
LOCUS
DEFINITION
UI-R-C2P-se-d-06-0-UI-r1 UI-R-C2P Rattus norvegicus cDNA clone
BF542321
UI-R-C2P-se-d-06-0-UI 5', mRNA sequence.

ACCESSION BF542321.1 GI:11633428
VERSION
KEYWORDS
SOURCE
ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE

JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 349)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares lab Clone distribution:
This clone is also available through Research Genetics (www.rsgen.com)
LNL (info@image.llnl.gov). IMAGE ID= 1768584
Seq primer: M13 Forward.

FEATURES

source

1. 349
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C2P-se-d-06-0-UI"
/clone_id="UI-R-C2P"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Bco RI; The UI-R-C2P
library is a subtracted library derived from the UI-R-C1
library, which is a subtracted library derived from the
UI-R-C0 library. The UI-R-C0 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture. The subtracted library (UI-R-C2P) was
constructed as follows: PCR amplified cDNA inserts from
UI-R-C1 clones from which 3' ESTs had been derived was
used as a driver in a hybridization with the UI-R-C1
library in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library) was
purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
UI-R-C2P library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996) "

BASE COUNT

121 a 88 c 67 g 73 t

Query Match 58.2%; Score 19.8; DB 12; Length 349;
Pred. No. 6.9e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 2 ACTGCTTACTTATGATCGTACTGTGACA 32

59 ACTGCTTACTTATGATCGTACTGTGACA 29

RESULT 8				
LOCUS	AL387696/c			
DEFINITION	AL387696	451 bp	mRNA	linear EST 03-AUG-2000
	MCB04409P1	McBC Medicago truncatula cDNA clone MCB044B09 T3,	451 bp	mRNA linear EST 03-AUG-2000
ACCESSION	AL387696			
VERSION	AL387696.1	GI:9687447		
KEYWORDS	EST.			
SOURCE	barrel medic.			
ORGANISM	Medicago truncatula			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	SpERMotophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	Rosidae; eurosoid I; Fabales; Fabaceae; Papilionoideae; Trifolieae;			
	Medicago.			
REFERENCE	1 (bases 1 to 451)			
AUTHORS	Journe'e,E.P., Crespeau,H., van-Tuijnen,D., Gouzy,J., Jailion,O., Nebel,A., Carreau,V., Chatagnier,O., Kahn,D., Glaninazzi-Pearson ,V. and Gamas,P.			
TITLE	Medicago truncatula ESTs from endomycorrhizal roots			
COMMENT	Unpublished (2000)			
CONTACT	Contact : Genoscope			
URL	Genoscope - Centre National de Sequencage			
	BP 191 91006 EVRY cedex - France			
	Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr			
	Contact : Pascal Gamas and Etienne-Pascal Journe'e, Laboratoire de			
	Biotique Molculaire des Relations Plantes-Microorganismes,			
	CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :			
	Mc-est@toulouse.inra.fr Website :			
	http://sequence.toulouse.inra.fr/McTruncatula.html).			

BASE COUNT	141 a	77 c	106 g	125 t	2 others
ORIGIN					
Query Match	58.2%; Score 19.8; DB 9; Length 451;				
Best Local Similarity	77.4%; Pred. No. 7,4e+02;				
Matches 24; Conservative	0; Mismatches 7; Indels 0; Gaps 0;				
QY	4 TCGCTACTTATCGATCGGTACTGCGAGACA 34				
Db	137 TCGCTCTTATCGATGAGGCTTGTGAATA 107				

LOCUS:	BM209528	502 bp	mRNA	linear	EST 31-JAN-2002
DEFINITION	C0649807.3 N1A Mouse Trophoblast Stem Cell cDNA Library (Long) Mus				
ACCESSION	musculus cDNA clone C0649807 3', mRNA sequence.				
VERSION	BM209528				
KEYWORDS	BM209528.1 GI:17766047				
SOURCE	EST.				
ORGANISM	house mouse.				
	Mus musculus				
	Eularyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 502)				
AUTHORS	Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Luo, A., Tanaka, T., Kunath, T., Rossant, J. and Ko, M.S.H.				
TITLE	Systematic Analyses of N1A Mouse Trophoblast Stem Cell cDNA Library (Long)				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Dawood B. Dudekula Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA Email: cdna@igsun.grc.nia.nih.gov Plate: C0649 row: B column: 07 Seq primer: -21M13 Forward High quality sequence stop: 502 POLYA=yes				
FEATURES	Location/Qualifiers				
Source	1..502				

	BASE COUNT	103 a	136 c	148 g	115 t
ORIGIN					
Query Match			58.2%	Score 19.8;	DB 13;
Best Local Similarity			77.4%;	Pred. No. 7.7e+02;	
Matches 24; Conservative			0; Mismatches	7; Indels	0; Gaps
0y	2	ACTCCTACTATGATCGGACGCTAGGAGA	32		
Db	487	ACCTCTCATTTATCTGCTGCTGCTAGAGA	457		

[illegible]

TITLE
JOURNAL
COMMENT

Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Shohiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. *Genome Res.* 10 (10), 1617-1630 (2000)
 wagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multichipillary sequencer. *Genome Res.*
 10 (11), 1751-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.

FEATURES
SOURCE

location/Qualifiers
 1..524
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370032F11"
 /clone_1lb="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"
 /note="pooled tissues; (tissue_type=cerebellum,
 dev_stage=16 days neonate, sex=mixed),
 (tissue_type=cerebellum, dev_stage=0 day neonate,
 sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
 sex=male), (tissue_type=whole body, dev_stages=days
 embryo, sex=mixed), (tissue_type=liver, dev_stage=13 days
 embryo, sex=mixed)"

BASE COUNT
 120 a 152 c 147 g 105 t

ORIGIN

Query Match 58.2% Score 19.8; DB 10; Length 524;
 Best Local Similarity 77.4%; Pred. No. 7.8e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

oy 2 ACTCCTACTTATGATCGGATGAGTATGAGA 32
 db 34 ACCTCTATTATTCGTCTGTCTGTGAGA 64

RESULT 12
 BB752196 530 bp mRNA linear EST 17-OCT-2001
 DEFINITION BB752196 RIKEN full-length enriched, melanocyte Mus musculus cDNA
 accession BB752196
 version BB752196.1 GI:16200663
 keywords EST.
 source house mouse.
 organism Mus musculus

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 530)
Akimura,T., Arakawa,T., Carinci,P., Furuno,M., Hangaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kondo,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toyota,T., Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Komoto,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagui,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Komoto,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
* Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES
source

Location/Qualifiers

1. 530
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G270012B20"
/clone_1ib="RIKEN full-length enriched, melanocyte"
/cell_type="melanocyte"
/note="pooled tissues; (tissue_type=cerebellum, dev_stage=16 days neonate, sex=mixed), (tissue_type=cerebellum, dev_stage=0 day neonate, sex=mixed), (tissue_type=hippocampus, dev_stage=adult, sex=male), (tissue_type=whole body, dev_stage=9 days embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days embryo, sex=mixed)"
embryo, sex=mixed"

BASE COUNT
ORIGIN

121 a 157 c 137 g 115 t
Query Match 58.2%; Score 19.8; DB 10; Length 530;
Best Local Similarity 77.4%; Pred. No. 7.8e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 ACTCTCTACTTATCGATCGTACTGTGAGA 32
Db 49 ACCCTCTATTATCTGTCTGTCTGTGAGA 79

RESULT 13
LOCUS

AZ331947/c 596 bp DNA linear GSS 29-SEP-2000

DEFINITION

IM060J06F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M006J06 F, DNA sequence.

ACCESSION
AZ331947
VERSION
AZ331947.1 GI:10959130
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Rellily and Wright,D., Weis,R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL
COMMENT

Unpublished (2000)
Contact: Robert B. Weis
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0060 row: J column: 06
Seq primer: CCGTGTAAACGACGCGCACT
Class: plasmid ends
High quality sequence stop: 596.

FEATURES
source

Location/Qualifiers

1. 596
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M006J06"
/clone_1ib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (GI:4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
ORIGIN

141 a 174 c 163 g 118 t
Query Match 58.2%; Score 19.8; DB 17; Length 596;
Best Local Similarity 77.4%; Pred. No. 8.1e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 ACTCTCTACTTATCGATCGTACTGTGAGA 32
Db 188 ACCCTCTATTATCTGTCTGTCTGTGAGA 158

RESULT 14
LOCUS

AZ831669 611 bp DNA linear GSS 20-FEB-2001

DEFINITION 2M011H02R Mouse 10kb plasmid UNGC1M library Mus musculus genomic
 ACCESSION clone UNGC2M011H02 R, DNA sequence.
 VERSION A2831669
 KEYWORDS A2831669.1 GI:13001577
 SOURCE GSS.
 ORGANISM house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 611)
 REFERENCE Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
 Jellam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
 M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,
 and Wright, D., Weis, R.
 Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weis
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0111 row: H column: 02
 Seq primer: CACGACGAAACAGCTATGAC
 Class: plasmid ends
 High quality sequence stop: 611.
 Location/Qualifiers
 1. 611
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UNG2M011H02"
 /clone_1b="Mouse 10kb plasmid UNGC1M library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, TI-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repeated with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 adaptor DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of pMD42 (g14732114|gb|AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptor complementary to the insert adaptor and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

BASE COUNT 153 a 123 c 182 g 153 t
 ORIGIN
 Query Match 58.2%; Score 19.8; DB 17; Length 611;
 Best Local Similarity 77.4%; Pred. No. 8.2e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 4 TCCTCTACTTATCGATCGTACTGTAGACA 34
 Db 252 TCCTCTACTTATCGATCGTACTGTAGACA 282

RESULT 15
 BM945933 615 bp mRNA linear EST 14-MAR-2002
 L6CUS

DEFINITION UI-M-EMO-bv1-f-05-0-UI.r1 NIH BMAP_EMO Mus musculus cDNA clone
 IMAGE:5692804 5', mRNA sequence.
 ACCESSION BM945933
 VERSION BM945933.1 GI:19429518
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 615)
 REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaabs@mail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 DNA Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: pyx-5.
 Location/Qualifiers
 1. 615
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5692804"
 /clone_1b="NIH BMAP EMO"
 /tissue_type="Whole brain"
 /dev_stage="embryo 18.5 dpc"
 /lab_host="DH10B (TI phage resistant)"
 /note="Organ: Brain; Vector: pyx-Aac; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured mRNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was size selected according to mRNA size fraction,
 ligated with EcoR I adaptor, digested with Not I, and then
 cloned directionally into pyx-Aac vector. The library tag
 sequence located between the Not I site and the polyA tail
 is CAGCAGGAC. This library was created for the
 University of Iowa Mouse Brain Molecular Anatomy Project
 (BMAP). 'Gene Discovery in the Developing Mouse Nervous
 System', supported by National Institute of Mental Health
 (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 122 a 178 c 175 g 140 t
 ORIGIN
 Query Match 58.2%; Score 19.8; DB 14; Length 615;
 Best Local Similarity 77.4%; Pred. No. 8.2e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 2 ACTCTCTACTTATCGATCGTACTGTAGACA 32
 Db 527 ACTCTCTACTTATCGTCTGTCTGTAGACA 557
 Search completed: May 11, 2003, 04:54:25
 Job time : 550.556 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic & nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 349.519 Seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-16

Perfect score: 33

Sequence: 1 gctctgtctactctctctctctctcaccagcca 33

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_ov:*

5: gb_pat:*

6: gb_ph:*

7: gb_pl:*

8: gb_pr:*

9: gb_ro:*

10: gb_ro:*

11: gb_ro:*

12: gb_ro:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_or:*

21: em_ov:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_ro:*

28: em_ro:*

29: em_ro:*

30: em_ro:*

31: em_ro:*

32: em_ro:*

33: em_ro:*

34: em_ro:*

35: em_ro:*

36: em_ro:*

37: em_ro:*

38: em_ro:*

39: em_ro:*

40: em_ro:*

41: em_ro:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	33	100.0	33	6	AX036750	AX036750 Sequence
2	24	72.7	180577	2	AC115145	AC115145 Rattus no
3	23.4	70.9	31069	2	AC130253	AC130253 Rattus no
4	23.4	70.9	147295	3	AGA439060	AJ439060 Anopheles
5	23.4	70.9	149788	8	ATT6621	AL589883 Arabidops
6	23.4	70.9	173535	2	AC021316	AC021316 Homo sapi
7	23.4	70.9	194974	2	AC107787	AC107787 Mus muscu
8	23.4	70.9	336062	2	AC091329	AC091329 Rattus no
9	23.4	70.9	345243	2	AC091347	AC091347 Rattus no
10	23.2	70.3	177899	10	AC123042	AC123042 Mus muscu
11	23	69.7	60921	9	AL391706	AL391706 Human DNA
12	23	69.7	224860	2	AC124546	AC124546 Mus muscu
13	23	69.7	229158	2	AC124707	AC124707 Mus muscu
14	22.6	68.5	234900	10	AL731717	AL731717 Mouse DNA
15	22.4	67.9	664	9	HS434005	AJ340055 Homo sapi
16	22.4	67.9	48042	2	AC020217	AC020217 Drosophila
17	22.4	67.9	83656	9	AC007401	AC007401 Homo sapi
18	22.4	67.9	103237	2	AC110484	AC110484 Rattus no
19	22.4	67.9	150765	2	AC094766	AC094766 Rattus no
20	22.4	67.9	167002	2	AC073041	AC073041 Homo sapi
21	22.4	67.9	169705	3	AC104704	AC104704 Drosophila
22	22.4	67.9	171907	2	AC116538	AC116538 Drosophila
23	22.4	67.9	181656	2	AC124900	AC124900 Rattus no
24	22.4	67.9	184620	2	AC096290	AC096290 Rattus no
25	22.4	67.9	197944	2	AC112748	AC112748 Rattus no
26	22.4	67.9	236303	3	AF168681	AF168681 Homo sapi
27	22.4	67.9	279679	3	AE003527	AE003527 Drosophila
28	22.2	67.3	1800	8	YSCMAT1A	L00060 Yeast (S.ce
29	22.2	67.3	2430	8	YSCMAT1A	L00313 Yeast (S.ce
30	22.2	67.3	2520	8	YSCMAT1A	L00059 Yeast (S.ce
31	22.2	67.3	2563	8	YSCMAT1A	X63853 S.cerevisia
32	22.2	67.3	9818	8	SCMAT1OC	248613 S.cerevisia
33	22.2	67.3	32481	8	SC8270	AL160262 Human DNA
34	22.2	67.3	53596	9	AL160262	Continuation (4 of
35	22.2	67.3	71804	2	AC099209_3	AC067861 Homo sapi
36	22.2	67.3	180978	2	AC067861	Continuation (3 of
37	22.2	67.3	190399	2	AC127960	AC127960 Rattus no
38	22.2	67.3	316613	8	SCCHR11I	X59720 S.cerevisia
39	22.2	66.7	1424	8	ASNRNVA	D28341 Aspergillus
40	22	66.7	70072	2	AC109210	AC109210 Mus muscu
41	22	66.7	100810	2	AC011991	AC011991 Homo sapi
42	22	66.7	110000	2	AC129176_2	Continuation (1 of
43	22	66.7	150980	2	AC107206	AC107206 Oryza sat
44	22	66.7	157250	2	AC027264	AC027264 Homo sapi
45	22	66.7	158770	9	AC007321	AC007321 Homo sapi

ALIGNMENTS

RESULT 1

AX036750 33 bp DNA

LOCUS AX036750

DEFINITION Sequence 16 from Patent WO0058485.

ACCESSION AX036750

VERSION AX036750.1 GI:11226259

KEYWORDS

ORGANISM

SOURCE

synthetic construct.

synthetic construct.

artificial sequences.

REFERENCE

1 (bases 1 to 33)

AUTHORS Rance, I., Theisen, M. and Gruber, V.

TITLE Chimera expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus

JOURNAL Patent: WO 0058485-A 16 05-OCT-2000;

Pred. No. is the number of results predicted by chance to have a

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
 ; GRUBER VERONIQUE (FR)
 LOCATION/Qualifiers
 1..33
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Guide deoxynucleotide building block G2"

BASE COUNT
 4 a 9 c 4 g 16 t

ORIGIN

Query Match 100.0%; Score 33; DB 6; Length 33;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GCTGTGTTACTTCTCTCTATTCAGCCA 33
 1 GCTGTGTTACTTCTCTCTATTCAGCCA 33

RESULT 2
 AC115145/c
 LOCUS
 DEFINITION
 AC115145 180577 bp DNA linear HTG 17-JUL-2002
 Rattus norvegicus clone CH230-201L12, *** SEQUENCING IN PROGRESS
 *** 53 unordered pieces.
 AC115145
 AC115145.2 GI:21743797
 HTG: HTGS PHASE1.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 180577)
 1 Mznu,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alebrooke,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Bileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
 Gabiela,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Hulvik,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Eten,C., Liu,J., Liu,W., Loulleged,H.,
 Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,B., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabac,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newson,J., Newton,S., Oguh,M., Okwundu,N.,
 Nguyen,N., Nickerson,E., Nwokemko,S., Oguh,M., Okwundu,N.,
 Ogunye,N., Oyedero,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,J.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
 Sodergren,E., Sonakke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tasey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Umani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Direct Submission
 Unpublished
 2 (bases 1 to 180577)
 Morley,K.C.
 Direct Submission
 Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 180577)
 Morley,K.C.
 Direct Submission
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 13, 2002 this sequence version replaced gi:19482203.

----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: GOB8
 Center clone name: CH230-201L12
 ----- Summary Statistics -----
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 132770 bases at least Q40
 Consensus quality: 139349 bases at least Q30
 Consensus quality: 144434 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 53 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1228: contig of 1228 bp in length
 * 1229 1338: gap of unknown length
 * 1329 2518: contig of 1190 bp in length
 * 2519 2618: gap of unknown length
 * 2619 3651: contig of 1033 bp in length
 * 3652 3751: gap of unknown length
 * 3752 5280: contig of 1533 bp in length
 * 5291 5390: gap of unknown length
 * 5391 6657: contig of 1267 bp in length
 * 6658 6757: gap of unknown length
 * 6758 7848: contig of 1091 bp in length
 * 7849 7948: gap of unknown length
 * 7949 9351: contig of 1403 bp in length
 * 9352 9451: gap of unknown length
 * 9452 10649: contig of 1198 bp in length
 * 10650 10749: gap of unknown length
 * 10750 12018: contig of 1269 bp in length
 * 12019 12118: gap of unknown length
 * 12119 13473: contig of 1355 bp in length
 * 13474 13573: gap of unknown length
 * 13574 15170: contig of 1597 bp in length
 * 15171 15270: gap of unknown length
 * 15271 16299: contig of 1029 bp in length
 * 16300 16399: gap of unknown length
 * 16400 17636: contig of 1237 bp in length
 * 17637 17737: gap of unknown length
 * 17737 19787: contig of 2051 bp in length
 * 19788 19887: gap of unknown length
 * 19888 21207: contig of 1320 bp in length
 * 21208 21308: gap of unknown length
 * 21308 22814: contig of 1507 bp in length
 * 22815 22914: gap of unknown length

22915 22915: contig of 1349 bp in length
 24264 24264: gap of unknown length
 25847 25847: contig of 1484 bp in length
 25848 25848: gap of unknown length
 27775 27775: contig of 1829 bp in length
 27777 27777: gap of unknown length
 30628 30628: contig of 2751 bp in length
 30727 30727: gap of unknown length
 32029 32029: contig of 1302 bp in length
 32030 32030: gap of unknown length
 32129 32129: gap of unknown length
 32130 32130: contig of 1660 bp in length
 33790 33790: gap of unknown length
 33885 33885: contig of 1496 bp in length
 35386 35386: gap of unknown length
 35486 35486: contig of 1634 bp in length
 37120 37120: gap of unknown length
 38423 38423: contig of 1204 bp in length
 38424 38424: gap of unknown length
 39854 39854: contig of 1462 bp in length
 39986 39986: gap of unknown length
 40086 40086: contig of 1895 bp in length
 41981 41981: gap of unknown length
 42080 42080: gap of unknown length
 45253 45253: contig of 3173 bp in length
 45254 45254: gap of unknown length
 45353 45353: gap of unknown length
 45354 45354: contig of 2959 bp in length
 48313 48313: gap of unknown length
 48412 48412: contig of 3431 bp in length
 51843 51843: gap of unknown length
 51943 51943: gap of unknown length
 55120 55120: contig of 3177 bp in length
 55121 55121: gap of unknown length
 55220 55220: gap of unknown length
 59133 59133: contig of 3913 bp in length
 59233 59233: gap of unknown length
 59234 59234: contig of 2329 bp in length
 61662 61662: gap of unknown length
 61663 61663: contig of 2625 bp in length
 64287 64287: gap of unknown length
 64288 64288: gap of unknown length
 66172 66172: contig of 1785 bp in length
 66173 66173: gap of unknown length
 66272 66272: gap of unknown length
 69183 69183: contig of 2911 bp in length
 69284 69284: gap of unknown length
 73838 73838: contig of 4555 bp in length
 73839 73839: gap of unknown length
 77332 77332: contig of 3794 bp in length
 77333 77333: gap of unknown length
 77832 77832: gap of unknown length
 80598 80598: contig of 2766 bp in length
 80699 80699: gap of unknown length
 84519 84519: contig of 3821 bp in length
 84520 84520: gap of unknown length
 88147 88147: contig of 3528 bp in length
 88247 88247: gap of unknown length
 88148 88148: gap of unknown length
 92434 92434: contig of 4187 bp in length
 92435 92435: gap of unknown length
 98849 98849: contig of 6315 bp in length
 98850 98850: gap of unknown length
 98949 98949: gap of unknown length
 10466 10466: contig of 5747 bp in length
 104796 104796: gap of unknown length
 110912 110912: contig of 6116 bp in length
 110913 110913: gap of unknown length
 118020 118020: contig of 7008 bp in length
 118120 118120: gap of unknown length
 124429 124429: contig of 6309 bp in length
 124529 124529: gap of unknown length
 130536 130536: contig of 6007 bp in length
 130635 130635: gap of unknown length
 139183 139183: contig of 8547 bp in length
 139283 139283: gap of unknown length
 147685 147685: contig of 8402 bp in length
 14785 14785: gap of unknown length
 156520 156520: contig of 8735 bp in length
 156521 156521: gap of unknown length
 156621 156621: gap of unknown length
 166173 166173: contig of 9553 bp in length
 166273 166273: gap of unknown length
 160577 160577: contig of 14304 bp in length.

FEATURES

Location/Qualifiers
 source 1. 180577

Query Match 72.7%; Score 24; DB 2; Length 180577;
 Best Local Similarity 84.4%; Pred. No. 64;

Matches 27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 CTGTTGTTACTTTCTTCTTATTCAGCCA 33
 Db 67869 CTGTTGTTACTTTCTTCTTATTCAGCCA 67838

RESULT 3

AC130253
 LOCUS

DEFINITION Rattus norvegicus clone CH230-50A6, *** SEQUENCING IN PROGRESS ***

AC130253 31069 bp DNA linear HTG 09-AUG-2002

20 unordered pieces.

AC130253 AC130253.1 GI:22164860

VERSION HTG; HTGS PHAS1.

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1. (bases 1 to 31069)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alshrocks,S.L., Anarlunga,H.C., Are,J.R., Ayale,M., Banke,T., Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonini,D., Bouck,J., Bowie,S., Brileva,M., Brown,B., Brown,M., Bryant,N.P., Bukey,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dedert,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabioli,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrill,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homes,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korrah,J., Kovar,C., Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louieged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Manua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mashiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G., Ogasunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pul,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Roite,M., Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umanai,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Woodson,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R.

TITLE

Unpublished

REFERENCE

2 (bases 1 to 31069)

AUTHORS

Worley,K.C.

TITLE

Direct Submission

Submitted (09-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

EHAGSSTGDKNLPAGTKILSYAEKASTNLVEKFTQSVRRIVQDVKDGSGPGMSRE
ELLETNERLEAVALRLEESYDTAKALVINLANKYGSKSQNTFNRYPMILKMIKXVI
RLQETQWTLVEIPEKQKLETPAPFVACISIMKSGSGGVTSAKLDEAERER
MERLEAMTTOIEOENTOMINDYRLIKKYTGARNIRLEKSEYGNKSYPIPRRTM
LKMDIMHDPMYMEVCHVDN"
640..654
/gene="4F11.9"
/number=1
655..1575
/gene="4F11.9"
/number=1
1576..1893
/gene="4F11.9"
/number=2
1894..2258
/gene="4F11.9"
/number=2
2259..3527
/gene="4F11.9"
/number=3
3528..4486
/gene="4F11.9"
/number=3
4487..4564
/gene="4F11.9"
/number=4
4565..4638
/gene="4F11.9"
/number=4
4639..4821
/gene="4F11.9"
/number=5
4822..4889
/gene="4F11.9"
/number=5
4890..5087
/gene="4F11.9"
/number=6
5088..5159
/gene="4F11.9"
/number=6
5160..5267
/gene="4F11.9"
/number=7
5268..5342
/gene="4F11.9"
/number=7
5343..5465
/gene="4F11.9"
/number=8
18352..21161
/gene="4F11.10"
join(18352..18654,21057..21161)
/gene="4F11.10"
join(18352..18654,21057..21161)
/gene="4F11.10"
/codon_start=1
/product="putative cytoskeletal regulator"
/protein_id="CAD27753.1"
/db_xref="GI:19577371"
/translation="MPFVKRVVTPKYIARSKLQPYAAPPIPVADNEFALTNVT
LSNALROLASLVYISNOIFTELHQLSIVNRSIGIKORIDRLSAKVDQDDPQVTVV
SSISRSRSTVYNLITCPSMAHAPGCRHT"
18352..18654
/gene="4F11.10"
/number=1
18655..21056
/gene="4F11.10"
/number=1
21057..21161
/gene="4F11.10"
/number=2
23341..31440

mRNA
CDS
/gene="11N17.1.1"
join(23341..23391,23609..23881,23940..23984,25261..25608,
25712..25945,26387..28585,28647..29090,30097..31440)
/gene="11N17.1.1"
join(23341..23391,23609..23881,23940..23984,25261..25608,
25712..25945,26387..28585,28647..29090,30097..31440)
/gene="11N17.1.1"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD27754.1"
/db_xref="GI:19577372"
/translation="MSQFQRLCEKCLAHGSGDLDTFVQIKHYTTKHVDGLFTVA
TSEVQVLEAQAAPVATIAEMDRIMINIGNSTEAPICTPVLSROKATAPITST
WTSKVLCEELPLPLRSGLVTIPTHSHPQPAVODLRMTSTALADVTGACSAK
IASHSSTNNVLPYITRESPTDLDAPTSNMAGDELDLDAITLAEIGIPSGNTIG
DNRVDHLPSPAOCRIILASKPAETIKIDTSGRAPDMSTKSLVCTYTTOKKGD
MAKPEEDYTRRRSLRRPRGRNRTIASDQRIEAVINKGPESTVEHYPCDGE
LMPAIRSKSGDILRKEPSPFAEMAKSFNMSPNSLKQGMKRLRMNRDSSASORD
KQODIDDFNIHDTTLAROPSVGSRRRGSDTDRERKLSHDKRPSYSSERSSTGILG
GTPAAYLPASINPVKLESTIRORRTALGNRDEPHSSGNNMSASGSGRTISGEIT
TNTNHPKSASSTLNHNPISSSAPPSIVRRRPNTSASSVTSEGITPDLOTF
DYHDEGGETSSVYSCDTEGVTSPFVHDSGLKTLKDEEPMPTLOSTTALSITSFSSG
NTTVSDVYVYKSGSTSTTSAGTICTYLAAGDSVSAASNLKPIPRKASITLKN
RNSVSNSTLERSTYSSTLSSTLSTGTIKNGVLIQKEVILKIQENNNRLDVAIG
NAPKVAAGTGAVOSSSEAVSSKLPPTAPEPESBESDVECVARTELKYKTTINTSRI
PMCIITPTNSDDEQPPAGSATAKAATORDNSTKAGIGAGKRRPTLTSTSPFELK
PDPRKHSLLPLANNVDRIGALPHLKSPTEBDATGFTTTSKDEBEAAGVGE
VYIRSKGRKTEMLTVRRRLATVLSGNINTEFEVYSINLPICNICEGANLALDCK
GVGVSEPAATELSBDALQSGQDGGSGPGLGGIGROALKGARVLDANGRAIF
SSDSLKRKRGAAHTTAPGPGCQOQOQVAGSAGGVSTTNVIGAPRLLPTRTLL
LKRPLVASRYGTPRIGPAPVAPAKTIVATILPNSAKPAQOPPPRLDARELSMRS
FSPYGADVSRGDRGGAAYAGAPLAAYQARSVAGTAAVAAGHAKLYPEGASHLD
RSIRNWVPDKLTVOGNPESATTLTPRPYARTVAPAGODVAPPALASAGFASPP
ASPIVDDTAPDPDPSGLYALPNKYKSGPLAGAVPKPMAGDVAIASEGSELPIPKCO
TNPTSTSKEDBALGALGPLHRRLLSNVNSLGNPSVHSGRSTREILBSSQPAAG
GTPRGHSVANSNVEVPNTCSRLGTPTKSLMDPFXLLAHGTKTHASPGSAMSAVEM
LKKSKESATRPKQDPGGSPASSGMALIDMSAPKMYFRIIAQGGYGSPTGA
GGGGGGGGGKGAAGRSNMYNNMRTGVSTALPEANSEEDIVPAPATATTTVSREE
PVASGKAPESSESGGGGVETISFKENIFLKEDENNFMGEVPRMTKSPAGFSTV
DGRKDMAGGRLLQARDOEADEBAASKQALETAL"
23341..23391
/gene="11N17.1.1"
/number=1
23392..23608
/gene="11N17.1.1"
/number=1
23609..23881
/gene="11N17.1.1"
/number=2
23882..23939
/gene="11N17.1.1"
/number=2
23940..23984
/gene="11N17.1.1"
/number=3
23985..25260
/gene="11N17.1.1"
/number=3
25261..25608
/gene="11N17.1.1"
/number=4
25609..25711
/gene="11N17.1.1"
/number=4
25712..25945
/gene="11N17.1.1"
Query Match 70.9%; Score 23.4; DB 3; Length 147295;
Best Local Similarity 81.8%; Pred No. 1,1e+02;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 GCTGTGTTACTTTCTTCTTATTCAGCCA 33
|||||

Db 97476 GCTGTGCTGCTTGCATTTCATTCACGCCA 97444

RESULT 5
ATT6G21/c 149788 bp DNA linear PLN 14-MAR-2001
LOCUS Arabidopsis thaliana DNA chromosome 5, BAC clone T6G21 (ESSA
project).
DEFINITION
ACCESSION
AL589883
VERSION
AL589883.1 GI:13374848
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 149788)
Bevan, M., Murphy, G., Ridley, P., Hudson, S., Bancroft, I., Mewes, H.W.,
Rudd, S., Lemcke, K. and Mayer, K.F.X.
Unpublished
2 (bases 1 to 149788)
EU Arabidopsis sequencing project.
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Direct Submission
Submitted (12-MAR-2001) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
Mayer@mips.gsf.de Project Coordinator: Mike Bevan, Molecular
Genetics Department, Cambridge Laboratory, John Innes Centre,
Colney Lane, NR4 7UJ Norwich UK, E-mail: michael.bevan@brc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.
FEATURES
source
location/Qualifiers
1. 149788
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
5957..8056
/gene="At5g21900"
join(5957..6637,6716..7152,7250..7411,7501..7576,
7657..7812,7934..8056)
/gene="At5g21900"
/note="similarity to DNA excision repair protein,
Schizosaccharomyces pombe, PIR:T41312
Contains Myb DNA-binding domain repeat signatures
AA502-510"
/codon_start=1
/product="DNA excision repair protein"
/protein_id="CAC34483.1"
/db_xref="GI:13374849"
/translation="MNSGDPNDRENRRINTDGEAIDALFMATEAVTRDPLSGPON
ODVMEERKSTADVTKGSDYWDYDEKRNFRKAKPALELHSDLKIVKTEBEVE
EVIKRDPEKSRRAKESIKGRROISTSDMEIDLNRRKAPSLVEISAVLAKNPLA
IKSLKLVFDRKRLKSLTVSGLGFDRRLMELLEDPSSEICANCVQVEDDVKIF
CDDDRVSLKVLIDLCGRSMTDYTIQFRRAPGFSPLTLLSQAFCLTDALLI
SKSPFLQYINLTGCSLLTYRALIILADKFGSTLIGSGCGIGKHGKFFSSLYEF
EKLNYSVAGVSVNDGVSRPFWRFSISLTDLSLANCVTDECMHIGRYCKLEA
LDITLDLKDLEFITEGCRVYKSLKLTNSRSDSCIAAPLEVSGGSIAREICINLV
RDVGPETAFSLAKYCKMLQFLDLSKCRKLKEDDRLRLRCCSLQSLKLTGMQVEET
YLERLSRSDVHITGLKLTSLYALHNDPSPVGAKEP"
5957..6637
/gene="At5g21900"
/number=1
6638..6715
/gene="At5g21900"
/number=1
6716..7152
/gene="At5g21900"
/number=2
7153..7249
/gene="At5g21900"
/number=2
7250..7411

introns
/gene="At5g21900"
/number=3
7412..7500
/gene="At5g21900"
/number=3
7501..7576
/gene="At5g21900"
/number=4
7577..7656
/gene="At5g21900"
/number=4
7657..7812
/gene="At5g21900"
/number=5
7813..7933
/gene="At5g21900"
/number=5
7934..8056
/gene="At5g21900"
/number=6
8057..8845
/gene="At5g21910"
8534..8845
/gene="At5g21910"
/codon_start=1
/product="putative protein"
/protein_id="CAC34484.1"
/db_xref="GI:13374850"
/translation="MKRFTTSKRLAISVLAVALSPYITDRLSEDEEELPSFM
FSPSLTLTLTALISTVSDSLTRPDYWIHRLCGSGGLVITLILVFLICAS
G"
8534..8845
/gene="At5g21910"
/number=1
9068..10483
/gene="At5g21920"
complement(join(9068..9314,9598..9800,10178..10483))
/gene="At5g21920"
complement(join(9068..9314,9598..9800,10178..10483))
/gene="At5g21920"
/note="similarity to various predicted proteins,
Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAC34485.1"
/db_xref="GI:13374851"
/translation="MEASANEPMKSLKSNPBGPIPNFVSLSSATQTPVLRNKNP
LTLPEVADSVKLIQDFHOSLSATEKFKGLHSLKNPLOEAVRLSSBPHLCDE
IRLRNTRVRPAMSNHNGFAVLPGDSVAGLVVANGLINLNTVITLVRLVLPSP
APPALVNPSTLCDDPYLNTFRGFIPPLGGLDLSPIIAFLVNAFTSSAMALPCELP
DGVSPASSETKWRRRRLSSHKDRPSSASMT"
complement(9068..9314)
/gene="At5g21920"
/number=1
complement(9315..9597)
/gene="At5g21920"
/number=1
complement(9598..9800)
/gene="At5g21920"
/number=2
complement(9801..10177)
/gene="At5g21920"
/number=2
complement(10178..10483)
/gene="At5g21920"
/number=3
10847..16439
/gene="At5g21930"
join(10847..10907,10989..11468,11567..11644,12017..12114,
13201..13254,13275..12461,12550..12714,13083..13193,
13335..13403,13695..13804,13981..14086,14288..14354,
14769..14848,15000..15629,15804..15935,16067..16217,
16348..16439)

/ncore="strong similarity to metal-transporting ATPase,
/ncore="At5g21930"
P-type, Arabidopsis thaliana, PIR:706003
contains Et-82 ATPase phosphorylation site AAs21-527"
/codon_start=1
/product="metal-transporting ATPase-like protein"
/protein_id="CAC34486.1"
/db_xref="GI:13374852"
/translation="MAMNLRPLPPSSLIHRIPISSNVEISSTQSFSTSSISYVS
ITSTPTPLDVSQMGCGCYARAKSYLMSDDVDAVAAYNMLETPLAKYKFEVYTD
TAEELARLTRESGEARVYSGVGLAENYKWKGVKSGEDLLVKSRRVAFTLVA
LCGSHSLIHSIGIHAQGIWDLNYSYKGLAAGALLGPRELIPFGIAPKPK
RSPMNNLVIGLSMAFSLISLVNPELEMPASFFDEYMLGVLIGRLERATL
QASPDNMRLESLSTOSRLVITSDNNTPEVDVLSDSICINAVSDIRVDSLIVP
GETPVYSGVGLQASVYDESLNGESLIPYKREGGSYAGITNMGGPLRIASSTG
STSKIVRWEDDAQNAAPQRLADIAGPYVYITMSLSMTFPAFWYVGHIPDVL
LNDIAGPDGMLALSLKAVDVLVYSCALSLATPTALIGTSLGARSGTILGGGV
LERASLIDCVALLKGTGLTEGRVYSGVASLGYEQEVLKMAAAVEKTAHPIDAKL
NEASLINAKTPERRGQITLEGPEFTLAIIDGRFVAAGSLSEWSDRLKKNSDMMV
SLLDHKSNTSSTSYRYKTVVYGRGEGSIIGALISDCLQDAFTVARKQEGIKT
VLSLDGDEGASVATYKAVNGISKSTYSLSPKKEFEFLNSGHRVAMVGDINDPA
PSLAQAVGIALKIQEENAAASPAVLVYRKLSHYVDALISQAQTSKYQNLAAVA
IANYVISTIPPAAGVLDPQIDFANTPSLSSGLMALSSIFVNSNLLQLHKSRTKNSL
"
10847..10907
/gene="At5g21930"
/number=1
10908..10988
/gene="At5g21930"
/number=1
10989..11468
/gene="At5g21930"
/number=2
11469..11566
/gene="At5g21930"
/number=2
11567..11644
/gene="At5g21930"
/number=3
11645..12016
/gene="At5g21930"
/number=3
12017..12114
/gene="At5g21930"
/number=4
12115..12200
/gene="At5g21930"
/number=4
12201..12254
/gene="At5g21930"
/number=5
12255..12374
/gene="At5g21930"
/number=5
12375..12461
/gene="At5g21930"
/number=5
12462..12549
/gene="At5g21930"

Query Match	23.4%	Score	DB 8	Length	149788;
Best Local Similarity	96.0%	Pred. No.	1,1e+02;		
Matches	24;	Conservative	0;	Mismatches	1;
				Indels	0;
				Gaps	0;
QY	1	GCATGTTACTTCTTCTCTCTAT	25		
DB	29248	GATGTTGTTACTTCTTCTCTAT	29224		

RESULT 6	
AC021316/c	
LOCUS	AC021316
DEFINITION	Homo sapiens chromosome 15, clone RP11-11J16, complete sequence.
	17535 bp DNA linear PRI 23-AUG-2002

ACCESSION	AC021316	
VERSION	AC021316.16	GI:22450641
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
1 (bases 1 to 173535)	Bliren, B., Nusbaum, C. and Lander, E.	Homo sapiens chromosome 15, clone RP11-11116	Unpublished	2 (bases 1 to 173535)	Bliren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

AUTHORS
Britten, E., Linton, U., Nussbaum, C., Lander, E., Aboian, H., Allen, N., Anderson, S., Baldwin, J., Barra, N., Beckery, R., Bedna, F., Boguslavsky, L., Bouhgalter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Domino, M., Doyle, M., Fenster, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McMan, P., McGurt, A., McKernan, K., McPheters, R., Melgrim, J., Menelus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliva, T. M., Peterson, K., Pierre, N., Pisan, C., Pollard, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teeleye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted 16-JAN-2000 Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 173535)
AUTHORS Birren, B., Linton, J., Nusbaum, C., Lande, E., Ali, A., Allen, N.,

Anderson, S., Bama, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collamore, A.,
Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,
Galgani, J., Gardyna, S., Ginde, S., Gird, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K.,
Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., MacDonald, P., Major, J., Margulis, N.,
Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C. H., O'Connor, F., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunhthong, P., Pletre, N.,
Pollara, V., Raymond, C., Retzl, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Saito, R., Schauer, S.,
Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strasser, N., Subramanian, A., Talamas, J., Testfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Triggilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (Pages 1 to 173535)

AUTHORS
 Batter, B., Nusbbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barua, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Ferro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Haags, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamauf, A.,
 Karates, A., Kells, C., Landers, T., Levine, R., Lindblad-Tsch, K.,
 Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Menges, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Notman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
 Phunkiang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schnupp, C., Retta, R., Seaman, S., Severy, P.,

TITLE
JOURNAL
COMMENT

Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tsefaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zember, I., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 330 Charles Street, Cambridge, MA 02141, USA
On Aug 23, 2002 this sequence version replaced gt:21307379.
All repeats were identified using RepeatMasker:
Smith, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L4301
Center clone name: 11_J_16

FEATURES

source
1. 173535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-11J16"
/clone_1kb="RRC1-11 Human Male BAC"
complement (1. 1513)
/rpt_family="L1"
repeat_region
1514. 1823
/rpt_family="AluY"
1824. 1848
/rpt_family="TNAn"
complement (1849. 2309)
/rpt_family="L1"
complement (2304. 3709)
/rpt_family="L1PBA"
3767. 3821
/rpt_family="TIGer5"
complement (4182. 4475)
/rpt_family="AluSc"
4497. 4517
/rpt_family="AT_rich"
complement (4523. 4779)
/rpt_family="TIGer3 (Golem)"
complement (4777. 4866)
/rpt_family="TIGer3 (Golem)"
complement (5350. 5474)
/rpt_family="L1ME"
complement (7827. 14234)
/rpt_family="L1PA7"
14502. 14555
/rpt_family="L2"
14893. 15214
/rpt_family="L1ME3"
15215. 15519
/rpt_family="Aluub"
15520. 16234
/rpt_family="L1ME3"
16247. 16848
/rpt_family="MLT1E2"
16863. 17022
/rpt_family="L1ME3"
17006. 17121
/rpt_family="L1MA4"
17239. 17445
/rpt_family="L2"
18801. 19118
/rpt_family="L1T855"
complement (19119. 19164)
/rpt_family="MLT2D"
19177. 19462
repeat_region

/rpt_family="TIGer3b"
19463. 19537
/rpt_family="TNAn"
complement (19538. 19841)
/rpt_family="AluY"
19842. 20687
/rpt_family="TIGer3b"
complement (20765. 21093)
/rpt_family="MLT2D"
21977. 22064
/rpt_family="CATATAAn"
complement (22066. 22279)
/rpt_family="TIGer3 (Golem)"
complement (22186. 22190)
/note="<30 qual SNGL region"
complement (22262. 22271)
/note="<30 qual SNGL region"
complement (22279. 22409)
/rpt_family="TIGer3 (Golem)"
22844. 23153
/rpt_family="Aluub"
23234. 23302
/rpt_family="CATATAAn"
23334. 23497
/rpt_family="MLT1J2"
complement (23589. 23593)
/note="<30 qual SNGL region"
complement (23619. 23623)
/note="<30 qual SNGL region"
23639. 23815
/rpt_family="MLT1J1"
complement (24116. 24458)
/rpt_family="L1ME3"
complement (24459. 24782)
/rpt_family="L2"
complement (24731. 24735)
/note="<30 qual SNGL region"
complement (24742. 24748)
/note="<30 qual SNGL region"
complement (24767. 24772)
/note="<30 qual SNGL region"
complement (24781. 24789)
/note="<30 qual SNGL region"
complement (24783. 25086)
/rpt_family="AluSg"
complement (24802. 24812)
/note="<30 qual SNGL region"
complement (24818. 24838)
/note="<30 qual SNGL region"
complement (24862. 24870)
/note="<30 qual SNGL region"
complement (25087. 25675)
repeat_region

Query Match 70.9%; Score 23.4; DB 9; Length 173535;
Best local similarity 81.8%; Pred. No. 1.1e+02;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 GCTGTTGTTACTTTCTTCTTCTATTCACCA 33
Db 80832 GCTGCTTAGACATTCTCTTCTATTCACCA 80800

RESULT 7
AC107787 194974 bp DNA linear HTG 12-APR-2002
LOCUS Mus musculus clone RP23-432P10, WORKING DRAFT SEQUENCE, 34 ordered
DEFINITION pieces.
ACCESSION AC107787
VERSION AC107787.2 GI:20143689
KEYWORDS HTG, HTGS_PHASE2, HTGS_DRAFT, HTGS_FULLTOP.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 194974)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-432p10
Unpublished
2 (bases 1 to 194974)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Darrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Margis, N., Mathews, C., McCarthy, M.,
McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Rettig, R., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 194974)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Darrellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margis, N., Mathews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Rettig, R., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (12-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 12, 2002 this sequence version replaced gi:18308489.
All repeats were identified using RepeatMasker:
Smith, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project information
Center project name: L20471
Center clone name: 432 P 10
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 185565 bases at least Q40
Consensus quality: 188852 bases at least Q30
Consensus quality: 190073 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 191674; sum-of-contigs
Quality coverage: 8.9 in Q20 bases; agarose-fp
Quality coverage: 9.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preferred.

1
1022 1121: contig of 1021 bp in length
1022 1121: gap of 100 bp
1122 1606: contig of 485 bp in length
1607 1706: gap of 100 bp
1707 2355: contig of 649 bp in length
2356 2455: gap of 100 bp
2456 3109: contig of 654 bp in length
3110 3209: gap of 100 bp
3210 4151: contig of 942 bp in length
4152 4251: gap of 100 bp
4252 4930: contig of 679 bp in length
4931 5030: gap of 100 bp
5031 6133: contig of 1093 bp in length
6124 6223: gap of 100 bp
6224 6989: contig of 766 bp in length
6990 7089: gap of 100 bp
7090 7786: contig of 707 bp in length
7797 7896: gap of 100 bp
7897 9525: contig of 1629 bp in length
9526 9625: gap of 100 bp
9626 10896: contig of 1271 bp in length
10897 10996: gap of 100 bp
10997 13026: contig of 2030 bp in length
13027 13126: gap of 100 bp
13127 45197: contig of 32071 bp in length
45198 45297: gap of 100 bp
45298 47519: contig of 2222 bp in length
47520 47619: gap of 100 bp
47620 49698: contig of 2079 bp in length
49699 49798: gap of 100 bp
49799 51950: contig of 2152 bp in length
51951 52050: gap of 100 bp
52051 54612: contig of 2562 bp in length
54613 54712: gap of 100 bp
54713 57085: contig of 2373 bp in length
57086 57185: gap of 100 bp
57186 60203: contig of 3018 bp in length
60204 60303: gap of 100 bp
60304 63082: contig of 2779 bp in length
63083 63182: gap of 100 bp
63183 66877: contig of 3695 bp in length
66878 66977: gap of 100 bp
66978 71786: contig of 4809 bp in length
71787 71886: gap of 100 bp
71887 77847: contig of 5961 bp in length
77848 77947: gap of 100 bp
77948 82080: contig of 4133 bp in length
82081 82180: gap of 100 bp
82181 88846: contig of 6666 bp in length
88847 88946: gap of 100 bp
88947 94785: contig of 5839 bp in length
94786 94885: gap of 100 bp
94886 102269: contig of 7384 bp in length
102270 102369: gap of 100 bp
102370 110158: contig of 7789 bp in length
110159 110258: gap of 100 bp

```

* 110259 120287: contig of 10029 bp in length
* 120288 120387: gap of 100 bp
* 120388 134724: contig of 14337 bp in length
* 134725 134824: gap of 100 bp
* 134825 151821: contig of 16997 bp in length
* 151822 151921: gap of 100 bp
* 151922 174022: contig of 22101 bp in length
* 174023 174122: gap of 100 bp
* 174123 193613: contig of 19490 bp in length
* 193613 193712: gap of 100 bp
* 193713 194974: contig of 1262 bp in length.
Location/Qualifiers
1. 194974
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-432P10"
/clone_1db="RP21-23 Female Mouse BAC"
1. 1021
/misc_feature
/notes="assembly_fragment
vector side:left"
1122_1606
/misc_feature
/notes="assembly_fragment"
1707_2355
/misc_feature
/notes="assembly_fragment"
2456_3109
/misc_feature
/notes="assembly_fragment"
3210_4151
/misc_feature
/notes="assembly_fragment"
4252_4930
/misc_feature
/notes="assembly_fragment"
5031_6123
/misc_feature
/notes="assembly_fragment"
6224_6989
/misc_feature
/notes="assembly_fragment"
7090_7796
/misc_feature
/notes="assembly_fragment"
7897_9525
/misc_feature
/notes="assembly_fragment"
9626_10896
/misc_feature
/notes="assembly_fragment"
10997_13026
/misc_feature
/notes="assembly_fragment"
13127_45197
/misc_feature
/notes="assembly_fragment"

Query Match 70.9%; Score 23.4; DB 2; Length 194974;
Best Local Similarity 81.8%; Pred. No. 1e+02;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTTGTTACTTCTCTCTATTCAGCA 33
Db 49919 GCTTTGTTATTTTCATCTTCATTCAGCAA 49951

RESULT 8
AC091229/c 336062 bp DNA linear HTG 31-JUN-2002
LOCUS Rattus norvegicus clone CH230-1A3, *** SEQUENCING IN PROGRESS ***
DEFINITION 185 unordered pieces.
AC091229
AC091229.11 GI:21953795
VERSION HTG; HTGS PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 336062)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alibrooke,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbada,J., Benton,J., Bimaga,K., Blahenbury,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,

```

```

Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.P., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlsen,E., Kelly,S., Khan,U., King,L., Korah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtenarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Meig,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Noser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokewo,S., Ogih,M., Okwundu,G.,
Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Stason,I.,
Sodergren,E., Sonatke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,D., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Uemami,K., Vaequez,L., Vera,V., Villalton,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 336062)
Worley,K.C.
Direct Submission
Submitted (05-APR-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 336062)
Worley,K.C.
Direct Submission
Submitted (31-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 24, 2002 this sequence version replaced gi:20976487.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: TUNJ
Center clone name: CH230-1A3
----- Summary Statistics
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-primer Body: 37% of reads
Chemistry: Dye-terminator Big Dye: 63% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 151384 bases at least Q40
Consensus quality: 169228 bases at least Q30
Consensus quality: 178295 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 185 contigs. The true order of the pieces

```

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1104: contig of 1104 bp in length
 * 1105 1204: gap of unknown length
 * 1205 2243: contig of 1039 bp in length
 * 2244 2343: gap of unknown length
 * 2344 3538: contig of 1195 bp in length
 * 3539 3638: gap of unknown length
 * 3639 4762: contig of 1124 bp in length
 * 4763 4862: gap of unknown length
 * 4863 5906: contig of 1044 bp in length
 * 5907 6006: gap of unknown length
 * 6007 7135: contig of 1129 bp in length
 * 7136 7235: gap of unknown length
 * 7236 8369: contig of 1134 bp in length
 * 8370 8469: gap of unknown length
 * 8470 9516: contig of 1047 bp in length
 * 9517 9616: gap of unknown length
 * 9617 10653: contig of 1037 bp in length
 * 10654 10753: gap of unknown length
 * 10754 11866: contig of 1113 bp in length
 * 11867 11966: gap of unknown length
 * 11967 13111: contig of 1145 bp in length
 * 13112 13211: gap of unknown length
 * 13212 14246: contig of 1035 bp in length
 * 14247 14346: gap of unknown length
 * 14347 15546: contig of 1200 bp in length
 * 15547 15646: gap of unknown length
 * 15647 16761: contig of 1115 bp in length
 * 16762 16861: gap of unknown length
 * 16862 17975: contig of 1114 bp in length
 * 17976 18075: gap of unknown length
 * 18076 19170: contig of 1095 bp in length
 * 19171 19270: gap of unknown length
 * 19271 20440: contig of 1170 bp in length
 * 20441 20540: gap of unknown length
 * 20541 21647: contig of 1107 bp in length
 * 21648 21747: gap of unknown length
 * 21748 22807: contig of 1060 bp in length
 * 22808 22907: gap of unknown length
 * 22908 24027: contig of 1120 bp in length
 * 24028 24127: gap of unknown length
 * 24128 25305: contig of 1178 bp in length
 * 25306 25405: gap of unknown length
 * 25406 26465: contig of 1060 bp in length
 * 26466 26565: gap of unknown length
 * 26566 27794: contig of 1229 bp in length
 * 27795 27894: gap of unknown length
 * 27895 28908: contig of 1014 bp in length
 * 28909 29008: gap of unknown length
 * 29009 30142: contig of 1134 bp in length
 * 30143 30242: gap of unknown length
 * 30243 31433: contig of 1191 bp in length
 * 31434 31533: gap of unknown length
 * 31534 32577: contig of 1044 bp in length
 * 32578 32677: gap of unknown length
 * 32679 33724: contig of 1047 bp in length
 * 33725 33824: gap of unknown length
 * 33825 34898: contig of 1074 bp in length
 * 34899 34998: gap of unknown length
 * 34999 36173: contig of 1175 bp in length
 * 36174 36273: gap of unknown length
 * 36274 37301: contig of 1028 bp in length
 * 37302 37401: gap of unknown length
 * 37402 38832: contig of 1431 bp in length
 * 38833 38932: gap of unknown length
 * 38933 40239: contig of 1307 bp in length
 * 40240 40339: gap of unknown length
 * 40340 41512: contig of 1173 bp in length

41513 41612: gap of unknown length
 * 41613 42959: contig of 1347 bp in length
 * 42960 43059: gap of unknown length
 * 43060 44103: contig of 1044 bp in length
 * 44104 44203: gap of unknown length
 * 44204 45329: contig of 1128 bp in length
 * 45330 45429: gap of unknown length
 * 45430 46682: contig of 1253 bp in length
 * 46683 46782: gap of unknown length
 * 46783 48017: contig of 1235 bp in length
 * 48018 48117: gap of unknown length
 * 48118 49331: contig of 1214 bp in length
 * 49332 49431: gap of unknown length
 * 49432 50543: contig of 1112 bp in length
 * 50544 52145: contig of 1502 bp in length
 * 52146 52246: gap of unknown length
 * 52247 53574: contig of 1329 bp in length
 * 53575 53674: gap of unknown length
 * 53675 55265: contig of 1591 bp in length
 * 55266 55365: gap of unknown length
 * 55366 56665: contig of 1300 bp in length
 * 56666 56765: gap of unknown length
 * 56766 57808: contig of 1043 bp in length
 * 57809 57908: gap of unknown length
 * 57909 58981: contig of 1073 bp in length
 * 58982 59081: gap of unknown length
 * 59082 60199: contig of 1117 bp in length
 * 60199 60299: gap of unknown length
 * 60299 61324: contig of 1025 bp in length
 * 61324 61423: gap of unknown length
 * 61424 63199: contig of 1776 bp in length
 * 63199 63299: gap of unknown length
 * 63300 64397: contig of 1099 bp in length
 * 64398 64497: gap of unknown length
 * 64498 65550: contig of 1053 bp in length
 * 65551 67161: gap of unknown length
 * 67161 67161: contig of 1511 bp in length

Query Match 70.9%; Score 23.4; DB 2; Length 336062;
 Best Local Similarity 81.8%; Pred. No. 96;
 Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 GCCTGTCTACTTTCTTCTTCTTATTCAGCA 33
 Db 245630 GCTCTTCTGCTCTCTTCTTCTTACAGCA 245598

RESULT 9

AC091347/c

LOCUS

DEFINITION

AC091347

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 345243)
 Muzny,D.M., Adams,C.C., Adio-oduola,B., Ali-osman,F.R., Allen,C.,
 Alibrooks,S.L., Amaralunga,H.C., Are,O.R., Ayale,M., Banks,T.,
 Barbarella,J., Benton,D., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowls,S., Brileva,M., Brown,M., Brown,M., Bryant,N.P.,
 Buhay,C., Burck,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,

	Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Baccotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabris, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gottell, J.H., Guvera, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, R., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homes, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C., Kratochvic, J., Kulesh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louieged, H., Lozad, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Mabeshwar, M., Mapa, P., Martin, R., Martindale, A., Martinez, B., Massey, E., Mahoney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Mosher, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nuyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okunom, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Severy, G., Scheerer, S., Scott, G., Shen, H., Shoshitari, N., Slason, I., Sodergren, E., Sonakke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansy, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Umami, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.
TITLE	Unpublished
JOURNAL	Direct Submission
REFERENCE	2 (bases 1 to 345243)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (18-APR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (bases 1 to 345243)
AUTHORS	Worley, K.C.
TITLE	Direct Submission
JOURNAL	Submitted (24-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT	On Jul 24, 2002 this sequence version replaced gi:20162995. ----- Genome Center ----- Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc.help@bcm.tmc.edu ----- Project Information ----- Center project name: TUD8 Center clone name: CH230-1B12 ----- Summary Statistics ----- Sequencing vector: Plasmid; Sequencing vector: M13; Chemistry: Dye-primer Bodipy; 33% of reads Chemistry: Dye-terminator Big Dye; 67% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 247030 bases at least Q40 Consensus quality: 258903 bases at least Q30 Consensus quality: 264786 bases at least Q20 ----- * NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html). * NOTE: This is a 'working draft' sequence. It currently * consists of 108 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. 1 1034: contig of 1034 bp in length 46442

* 46542 47953: contig of 1412 bp in length
 * 47954 48053: gap of unknown length
 * 48054 49285: contig of 1232 bp in length
 * 49286 49385: gap of unknown length
 * 49386 50659: contig of 1274 bp in length
 * 50660 50759: gap of unknown length
 * 50760 52545: contig of 1786 bp in length
 * 52546 52645: gap of unknown length
 * 52646 53911: contig of 1266 bp in length
 * 53912 54011: gap of unknown length
 * 54012 55564: contig of 1553 bp in length
 * 55565 55664: gap of unknown length
 * 55665 56920: contig of 1256 bp in length
 * 56921 57020: gap of unknown length
 * 57021 58310: contig of 1290 bp in length
 * 58311 58410: gap of unknown length
 * 58411 59529: contig of 1119 bp in length
 * 59530 59629: gap of unknown length
 * 59630 61186: contig of 1557 bp in length
 * 61187 61287: gap of unknown length
 * 61287 62625: contig of 1339 bp in length
 * 62626 62725: gap of unknown length
 * 62726 64167: contig of 1442 bp in length
 * 64168 64267: gap of unknown length
 * 64268 66271: contig of 2004 bp in length
 * 66272 66371: gap of unknown length
 * 66372 68090: contig of 1719 bp in length
 * 68091 68190: gap of unknown length
 * 68191 70130: contig of 1940 bp in length
 * 70131 70230: gap of unknown length
 * 70231 71628: contig of 1398 bp in length

Query Match 70.3%; Score 23.4; DB 2; Length 345243;
 Best Local Similarity 81.8%; Pred. No. 96;
 Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTTTACTTCTTCTTCTATTCAGCACA 33
 Db 133398 GCTGTGTCGTCTCTCTTCTTATACAGCACA 133366

RESULT 10
 AC123042/c 177899 bp DNA linear ROD 13-AUG-2002
 LOCUS Mus musculus chromosome 10 clone RP24-545017, complete sequence.
 DEFINITION AC123042
 ACCESSION AC123042 GI:22213542
 VERSION AC123042.2
 KEYWORDS HTG.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 177899)
 McPherson, J.D. and Waterston, R.H.
 TITLE The sequence of Mus musculus clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 177899)
 AUTHORS McPherson, J.D. and Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 3 (bases 1 to 177899)
 McPherson, J.D. and Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (13-AUG-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 COMMENT On Aug 13, 2002 this sequence version replaced gi:21217585.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc/index.shtml>
 Contact: submissions@wustl.edu

----- Project Information -----
 Center project name: M BB0545017

 FEATURES
 source
 1. 177899
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="10"
 /clone="RP24-545017"

BASE COUNT 52391 a 32822 c 33924 g 58762 t
 ORIGIN

Query Match 70.3%; Score 23.2; DB 10; Length 177899;
 Best Local Similarity 89.3%; Pred. No. 1.2e+02;
 Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTGTTTACTTCTTCTTCTTCTATTCACA 29
 Db 144939 CTGTTGTTCTTTCTTCTTCTTCACTCCA 144912

RESULT 11
 AL391706/c 60921 bp DNA linear PRI 15-NOV-2001
 LOCUS Human DNA sequence from clone RP11-431K17 on chromosome 1, complete
 DEFINITION sequence.
 ACCESSION AL391706
 VERSION AL391706.9 GI:16972964
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 60921)

REFERENCE 1 Hammond, S.
 AUTHORS Direct Submission
 TITLE Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
 JOURNAL Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridge, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 On Nov 16, 2001 this sequence version replaced gi:15795440.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; SW:
 SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
 RP11-431K17 is from the library RP11-11.2 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone
 RP11-431K17. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP11-445P19 is at 58922 in this
 sequence. The true right end of clone RP11-108M21 is at 2000 in
 this sequence.

FEATURES
 Location/Qualifiers


```

source
1. .60921
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-431K17"
/clone_1b="RP11-11.2"
BASE COUNT      19409 a 11328 c 10604 g 19580 t
ORIGIN

Query Match      69.7%; Score 23; DB 9; Length 60921;
Best Local Similarity 83.9%; Pred. No. 1.7e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

3 TGTGTACTTTCTCTCTATTCACGCCA 33
Db 38675 TGTCTGACTTTCTCTCTATTCACGCCA 38645

RESULT 12
AC124546      224860 bp      DNA      linear      HTG 11-AUG-2002
LOCUS
DEFINITION
Mus musculus chromosome UNK clone RP23-273G17, WORKING DRAFT
SEQUENCE, 2 unordered pieces.
AC124546
AC124546.2 GI:22203919
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 224860)
McPherson,J.D. and Waterston,R.H.
TITLE
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 224860)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (14-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 224860)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (11-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Aug 11, 2002 this sequence version replaced gi:21426675.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BA0273G17
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 223625 bases at least Q40
Consensus quality: 224098 bases at least Q30
Consensus quality: 224340 bases at least Q20
Insert size: 197000; agarose-fp
Insert size: 224760; sum-of-contigs
Quality coverage: 11.52 in Q20 bases; agarose-fp
Quality coverage: 11.90 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 7505: contig of 7505 bp in length
* 7506 7605: gap of unknown length
* 7606 224860: contig of 217255 bp in length.
Location/Qualifiers
1. .224860
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-273G17"
1. .7505
/clone="assembly_name:Contigs5"
/note="assembly_name:Contigs5"
misc_feature 7606. .224860
misc_feature /note="assembly_name:Contigs6"
BASE COUNT      72803 a 40558 c 41158 g 70219 t      122 others
ORIGIN

Query Match      69.7%; Score 23; DB 2; Length 224860;
Best Local Similarity 83.9%; Pred. No. 1.4e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

2 CTGTGTACTTTCTCTCTATTCACGCC 32
Db 211703 CTGTGTGCTTTCTCTCTATTCACGCC 211733

RESULT 13
AC124707      291588 bp      DNA      linear      HTG 22-JUN-2002
LOCUS
DEFINITION
Mus musculus chromosome UNK clone RP24-72H16, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC124707
AC124707.1 GI:21427964
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 291588)
McPherson,J.D. and Waterston,R.H.
1 (bases 1 to 291588)
McPherson,J.D. and Waterston,R.H.
TITLE
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 291588)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (15-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 291588)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (22-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BB0073H16
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 286901 bases at least Q40
Consensus quality: 288260 bases at least Q30
Consensus quality: 288883 bases at least Q20
Insert size: 197000; agarose-fp

```


Insert size: 290988; sum-of-contigs
Quality coverage: 15.82 in Q20 bases; agarose-fp
Quality coverage: 9.51 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 15773: contig of 15773 bp in length
* 15774 15873: gap of unknown length
* 15874 41215: contig of 25342 bp in length
* 41216 59868: gap of unknown length
* 59868 59969: gap of unknown length
* 59969 93088: contig of 33119 bp in length
* 93088 128533: gap of unknown length
* 128533 128634: gap of unknown length
* 128634 194926: contig of 66293 bp in length
* 194927 195026: gap of unknown length
* 195027 291588: contig of 96562 bp in length.

FEATURES

source

1. 291588
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP24-73H16"

misc_feature
/note="assembly_name:Contig12"
15874..41215
/note="assembly_name:Contig13"
41316..59868
/note="assembly_name:Contig14"
59969..93087
/note="assembly_name:Contig15"
93188..128533
/note="assembly_name:Contig16"
128634..194926
/note="assembly_name:Contig17"
195027..291588
/note="assembly_name:Contig18"
BASE COUNT 92191 a 54654 c 54286 g 89797 t 660 others
ORIGIN

Query Match 69.7%; Score 23; DB 2; Length 291588;
Best Local Similarity 83.9%; Pred. No. 1.4e+02;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTCTTCTTCTATTCAGCC 32

Db 141855 CTGTTTGTCTTCTCTCTATTCACCC 141825

RESULT 14

AL731717 234900 bp DNA linear ROD 24-JUL-2002
LOCUS Mouse DNA sequence from clone RP23-92C4 on chromosome X, complete
DEFINITION
ACCESSION AL731717
VERSION AL731717.10 GI:21955600
KEYWORDS
SOURCE
ORGANISM

Mus musculus.
Mus musculus.
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 234900)
AUTHORS Brown, J.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,

COMMENT

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On July 25, 2002 this sequence version replaced gi:21690011.

Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at:
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-92C4 is
from the RP23-Mouse PAC library
constructed by the group of Pletier de Jong.
For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR::PBACE3.6.

FEATURES

source

1. 234900
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP23-92C4"
/clone_1b="RP23-92C4"

BASE COUNT 72446 a 45634 c 45215 g 71605 t
ORIGIN

Query Match 68.5%; Score 22.6; DB 10; Length 234900;
Best Local Similarity 86.2%; Pred. No. 2e+02;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCTGTTGTTACTTTCTTCTTCTATTCGA 29

Db 48695 GCTGTTGTTACTTTGCTTCTGTTACA 48723

RESULT 15

HSA340055/c 664 bp DNA linear PRI 18-JUL-2002
LOCUS Homo sapiens genomic sequence surrounding Not1 site, clone
DEFINITION
ACCESSION HSA340055
VERSION HSA340055
KEYWORDS
SOURCE
ORGANISM

Homo sapiens.
Homo sapiens.
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 664)
AUTHORS Kutsenko, A.S., Gizetullin, R.Z., Al-Amin, A.N., Wang, F., Kvaeha, S.M.,
Podowski, R.M., Matushkin, Y.G., Gyanchandani, A., Muzvenko, O.V.,
Levitsky, V.G., Kolchanov, N.A., Protopopov, A.I., Kashuba, V.I.,
Kiselev, L.L., Wasserman, W., Wahlstedt, C., and Zabarovsky, B.R.
TITLE Not1 flanking sequences: a tool for gene discovery and verification
of the human genome
JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)
MEDLINE 22131767
PUBMED 12136098
REFERENCE 2 (bases 1 to 664)

AUTHORS Zabarcovsky, B.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
 Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
 Sweden

FEATURES
 SOURCE Location/Qualifiers
 1. .664
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NR5-1122C"

BASE COUNT 231 a 144 c 148 g 141 t
 ORIGIN

Query Match 67.9%; Score 22.4; DB 9; Length 664;
 Best Local Similarity 81.2%; Pred. No. 5.3e+02;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTTCTTCTTATTCAGCCA 33
 |||||
 Db 419 CTGTTGTTAGTTTCTTATTCGTTTCAGTCA 368

Search completed: May 11, 2003, 06:12:35
 Job time : 604.519 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 64.1805 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-16

Perfect score: 33
Sequence: 1 gcctgtctactctctctctctcaccagcca 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq.101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	100.0	33	21	AAA96850
2	22.4	67.9	569	22	AAK72229
3	22.4	67.9	47319	22	AAK4813
4	22.4	67.9	47319	22	AAK72230
5	22.4	67.9	236303	22	AAK11614
6	22	66.7	458	21	AAK43567
7	22	66.7	13079	22	ABAI8638
8	22	66.7	25967	22	ABAI8638
9	21.6	65.5	144	22	ABAI75350

10	21.6	65.5	144	22	ABA39992
11	21.6	65.5	144	22	AAK23906
12	21.6	65.5	144	22	AAK49980
13	21.6	65.5	144	22	AAI27077
14	21.6	65.5	568	22	ABA62842
15	21.6	65.5	568	22	ABA30124
16	21.6	65.5	568	22	AAK1243
17	21.6	65.5	568	22	AAK37042
18	21.6	65.5	568	22	AAI17884
19	21.6	65.5	1057	22	ABK35902
20	21.4	64.8	5322	22	AAI37484
21	21.4	64.8	10644	23	ABAI03638
22	21.4	64.8	30568	22	AAI37486
23	21.2	64.2	242	20	AAV89996
24	21.2	64.2	1904	22	AAK05148
25	21.2	64.2	2720	24	ABK10802
26	21.2	64.2	3699	24	ABK10805
27	20.8	63.0	222	22	AAK17163
28	20.8	63.0	1251	24	ABK72973
29	20.8	63.0	2312	22	AAK70886
30	20.8	63.0	2745	24	ABK84523
31	20.6	62.4	2823	21	AAK52632
32	20.6	62.4	2823	23	ABK25094
33	20.4	61.8	170	15	AAK76837
34	20.4	61.8	1224	23	AAK87626
35	20.4	61.8	1251	22	ABAI07827
36	20.4	61.8	1251	22	AAI03631
37	20.4	61.8	112190	22	AAH44801
38	20.2	61.2	693	21	AAI16256
39	20.2	61.2	1140	21	AAK90085
40	20.2	61.2	1286	23	AAK85199
41	20.2	61.2	1301	24	ABN95636
42	20.2	61.2	3758	17	AAK44318
43	20.2	61.2	4037	24	ABN97447
44	20.2	61.2	4037	24	ABK63716
45	20.2	61.2	4959	22	AAH81756

ALIGNMENTS

RESULT 1	
AAA96850	
ID	AAA96850 standard; DNA; 33 BP.
XX	XX
AC	AAA96850;
XX	XX
DT	19-FEB-2001 (first entry)
XX	XX
DE	Guide desoxynucleotide building block G2.
XX	XX
KM	Promoter; intergenic region; Comelina yellow mottle virus;
KW	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
XX	transgenic plant; ss.
XX	XX
OS	Synthetic.
XX	XX
PN	WO200058485-A1.
XX	XX
PD	05-OCT-2000.
XX	XX
PF	29-MAR-2000; 2000WO-IB00370.
XX	XX
PR	29-MAR-1999; 99FR-0003925.
XX	XX
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	XX
PI	Rance I, Gruber V, Theisen M;
DR	WPI; 2000-647238/62.
XX	XX
PT	Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Disclosure; Page 24; 91pp; English.
XX
CC The present sequence represents a guide deoxynucleotide building
CC block, which was used to link directional deoxynucleotide building
CC blocks during construction of chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
CC
SO Sequence 33 BP; 4 A; 9 C; 4 G; 16 T; 0 other;

Query Match 100.0%; Score 33; DB 21; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCTGTTACTTCTCTCTATTCAGCA 33
Db 1 GCTGTTACTTCTCTCTATTCAGCA 33

RESULT 2
AAK72229/c
ID AAK72229 standard; DNA; 569 BP.
XX
AC AAK72229;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27041.
XX
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosolic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN MO200157182-A2.
PD
XX 09-AUG-2001.
PD
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225277.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226682.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0230439.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234977.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236379.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240360.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241185.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.

XX	PR	08-NOV-2000;	2000US-0246525.
CC	PR	08-NOV-2000;	2000US-0246525.
XX	PR	08-NOV-2000;	2000US-0246526.
CC	PR	08-NOV-2000;	2000US-0246526.
XX	PR	08-NOV-2000;	2000US-0246527.
CC	PR	08-NOV-2000;	2000US-0246528.
XX	PR	08-NOV-2000;	2000US-0246532.
CC	PR	08-NOV-2000;	2000US-0246539.
XX	PR	08-NOV-2000;	2000US-0246609.
CC	PR	08-NOV-2000;	2000US-0246610.
XX	PR	08-NOV-2000;	2000US-0246611.
CC	PR	08-NOV-2000;	2000US-0246611.
XX	PR	08-NOV-2000;	2000US-0246613.
CC	PR	17-NOV-2000;	2000US-0249207.
XX	PR	17-NOV-2000;	2000US-0249208.
CC	PR	17-NOV-2000;	2000US-0249209.
XX	PR	17-NOV-2000;	2000US-0249210.
CC	PR	17-NOV-2000;	2000US-0249211.
XX	PR	17-NOV-2000;	2000US-0249212.
CC	PR	17-NOV-2000;	2000US-0249213.
XX	PR	17-NOV-2000;	2000US-0249214.
CC	PR	17-NOV-2000;	2000US-0249215.
XX	PR	17-NOV-2000;	2000US-0249264.
CC	PR	17-NOV-2000;	2000US-0249265.
XX	PR	17-NOV-2000;	2000US-0249297.
CC	PR	17-NOV-2000;	2000US-0249299.
XX	PR	17-NOV-2000;	2000US-0249300.
CC	PR	01-DEC-2000;	2000US-0250160.
XX	PR	01-DEC-2000;	2000US-0250391.
CC	PR	05-DEC-2000;	2000US-0251030.
XX	PR	05-DEC-2000;	2000US-0251988.
CC	PR	05-DEC-2000;	2000US-0256719.
XX	PR	06-DEC-2000;	2000US-0251479.
CC	PR	08-DEC-2000;	2000US-0251856.
XX	PR	08-DEC-2000;	2000US-0251868.
CC	PR	08-DEC-2000;	2000US-0251869.
XX	PR	08-DEC-2000;	2000US-0251989.
CC	PR	08-DEC-2000;	2000US-0251990.
XX	PR	11-DEC-2000;	2000US-0254097.
CC	PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Barash SC, Ruben SM;		
XX	WPI; 2001-483426/52.		
XX	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,		
XX	useful for preventing, diagnosing and/or treating cancers and		
XX	metastasis -		
XX	Disclosure; SEQ ID NO 27041; 3071pp + Sequence Listing; English.		
XX	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)		
CC	antino acid sequences given in AAM82170 to AAM91921. (I) have cytostratic		
CC	activity, and can be used in gene therapy and vaccine production. (I)		
CC	proteins and polynucleotides may be used in the prevention, diagnosis and		
CC	treatment of diseases associated with inappropriate (I) expression. For		
CC	example, they may be used to treat disorders associated with decreased		
CC	expression by rectifying mutations or deletions in a patient's genome		
CC	that affect the activity of (I) by expressing inactive proteins or to		
CC	supplement the patients own production of (I). Additionally, (I)		
CC	polynucleotides may be used to produce the secreted (I), by inserting		
CC	the nucleic acids into a host cell and culturing the cell to express the		
CC	protein. (I) proteins and polynucleotides may be used to prevent,		
CC	diagnose and treat immune/hematopoietic-related diseases, especially		
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703		
CC	to AAK87694 represent human immune/hematopoietic antigen genomic		
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169		
CC	represent sequences used in the exemplification of the present invention.		
XX			

Seq	Sequence	569 BP; 162 A; 83 C; 156 G; 168 T; 0 other;
Query Match		67.9%; Score 22.4; DB 22; Length 569;
Best Local Similarity		81.2%; Pred. No. 91;
Matches	26; Conservative	0; Mismatches 6; Indels 0; Gaps 0;
Db	1 GCTGTGTTACTTCTTCTTCATTCGAGCC 32 551 GCTGATGCTACTTACTTCTTGTATTCACAC 520	
RESULT 3		
ID	AAK64813/c	
AC	AAK64813 standard; DNA; 47319 BP.	
XX	AAK64813;	
DT	06-NOV-2001 (first entry)	
XX		
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:19625.	
XX		
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KM	cytostatic; gene therapy; vaccine; metastasis; ds.	
XX		
OS	Homo sapiens.	
XX		
FN	W0200157182-A2.	
PD		
XX	09-AUG-2001.	
PF	17-JAN-2001; 2001WO-US01354.	
XX		
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	
PR	07-JUN-2000; 2000US-0209467.	
PR	28-JUN-2000; 2000US-0214886.	
PR	30-JUN-2000; 2000US-0215135.	
PR	07-JUL-2000; 2000US-0216647.	
PR	07-JUL-2000; 2000US-0216880.	
PR	11-JUL-2000; 2000US-0217487.	
PR	11-JUL-2000; 2000US-0217496.	
PR	14-JUL-2000; 2000US-0218290.	
PR	26-JUL-2000; 2000US-0220963.	
PR	26-JUL-2000; 2000US-0220964.	
PR	14-AUG-2000; 2000US-0224518.	
PR	14-AUG-2000; 2000US-0224519.	
PR	14-AUG-2000; 2000US-0225213.	
PR	14-AUG-2000; 2000US-0225214.	
PR	14-AUG-2000; 2000US-0225266.	
PR	14-AUG-2000; 2000US-0225267.	
PR	14-AUG-2000; 2000US-0225268.	
PR	14-AUG-2000; 2000US-0225270.	
PR	14-AUG-2000; 2000US-0225447.	
PR	14-AUG-2000; 2000US-0225757.	
PR	14-AUG-2000; 2000US-0225758.	
PR	14-AUG-2000; 2000US-0225759.	
PR	18-AUG-2000; 2000US-0226279.	
PR	22-AUG-2000; 2000US-0226681.	
PR	22-AUG-2000; 2000US-0226868.	
PR	22-AUG-2000; 2000US-0227182.	
PR	23-AUG-2000; 2000US-0227009.	
PR	30-AUG-2000; 2000US-0228924.	
PR	01-SEP-2000; 2000US-0229287.	
PR	01-SEP-2000; 2000US-0229343.	
PR	01-SEP-2000; 2000US-0229344.	
PR	01-SEP-2000; 2000US-0229345.	
PR	05-SEP-2000; 2000US-0229509.	

PR	05-SEP-2000	2000US-0229531	
PR	06-SEP-2000	2000US-0230437	
PR	06-SEP-2000	2000US-0230438	
PR	08-SEP-2000	2000US-0231242	
PR	08-SEP-2000	2000US-0231243	
PR	08-SEP-2000	2000US-0231244	
PR	08-SEP-2000	2000US-0231244	
PR	08-SEP-2000	2000US-0231413	
PR	08-SEP-2000	2000US-0231414	
PR	08-SEP-2000	2000US-0233080	
PR	08-SEP-2000	2000US-0233081	
PR	12-SEP-2000	2000US-0233166	
PR	14-SEP-2000	2000US-0233397	
PR	14-SEP-2000	2000US-0233398	
PR	14-SEP-2000	2000US-0233399	
PR	14-SEP-2000	2000US-0233400	
PR	14-SEP-2000	2000US-0233401	
PR	14-SEP-2000	2000US-0233063	
PR	14-SEP-2000	2000US-0233064	
PR	14-SEP-2000	2000US-0233065	
PR	21-SEP-2000	2000US-0234223	
PR	21-SEP-2000	2000US-0234274	
PR	25-SEP-2000	2000US-0234997	
PR	25-SEP-2000	2000US-0234997	
PR	26-SEP-2000	2000US-0235488	
PR	27-SEP-2000	2000US-0235834	
PR	27-SEP-2000	2000US-0235836	
PR	29-SEP-2000	2000US-0236327	
PR	29-SEP-2000	2000US-0236367	
PR	29-SEP-2000	2000US-0236367	
PR	29-SEP-2000	2000US-0236368	
PR	29-SEP-2000	2000US-0236369	
PR	29-SEP-2000	2000US-0236370	
PR	02-OCT-2000	2000US-0236802	
PR	02-OCT-2000	2000US-0237037	
PR	02-OCT-2000	2000US-0237037	
PR	02-OCT-2000	2000US-0237038	
PR	02-OCT-2000	2000US-0237039	
PR	12-OCT-2000	2000US-0237040	
PR	13-OCT-2000	2000US-0239935	
PR	13-OCT-2000	2000US-0239937	
PR	20-OCT-2000	2000US-0240960	
PR	20-OCT-2000	2000US-0241221	
PR	20-OCT-2000	2000US-0241785	
PR	20-OCT-2000	2000US-0241786	
PR	20-OCT-2000	2000US-0241787	
PR	20-OCT-2000	2000US-0241808	
PR	20-OCT-2000	2000US-0241809	
PR	20-OCT-2000	2000US-0241826	
PR	01-NOV-2000	2000US-0244617	
PR	08-NOV-2000	2000US-0246474	
PR	08-NOV-2000	2000US-0246475	
PR	08-NOV-2000	2000US-0246476	
PR	08-NOV-2000	2000US-0246476	
PR	08-NOV-2000	2000US-0246477	
PR	08-NOV-2000	2000US-0246478	
PR	08-NOV-2000	2000US-0246523	
PR	08-NOV-2000	2000US-0246524	
PR	08-NOV-2000	2000US-0246609	
PR	08-NOV-2000	2000US-0246610	
PR	08-NOV-2000	2000US-0246611	
PR	08-NOV-2000	2000US-0246613	
PR	17-NOV-2000	2000US-0249207	
PR	17-NOV-2000	2000US-0249208	
PR	17-NOV-2000	2000US-0249209	
PR	17-NOV-2000	2000US-0249210	
PR	17-NOV-2000	2000US-0249211	
PR	17-NOV-2000	2000US-0249212	
PR	17-NOV-2000	2000US-0249213	
PR	17-NOV-2000	2000US-0249214	
PR	17-NOV-2000	2000US-0249215	
PR	17-NOV-2000	2000US-0249216	

[illegible]

XX	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer
KM	Cytotoxic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
FN	MO200157182-A2.
XX	
XX	09-AUG-2001.
XX	
XX	17-JAN-2001; 2001WO-US01354.
XX	
FR	31-JAN-2000; 2000US-0179065.
FR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218230.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225256.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225470.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226868.
PR	23-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228294.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.
PR	01-SEP-2000; 2000US-0229344.
PR	01-SEP-2000; 2000US-0229345.
PR	05-SEP-2000; 2000US-0229509.
PR	05-SEP-2000; 2000US-0229513.
PR	06-SEP-2000; 2000US-0230437.
PR	06-SEP-2000; 2000US-0230438.
PR	08-SEP-2000; 2000US-0231242.
PR	08-SEP-2000; 2000US-0231243.
PR	08-SEP-2000; 2000US-0231244.
PR	08-SEP-2000; 2000US-0231413.
PR	08-SEP-2000; 2000US-0231414.
PR	08-SEP-2000; 2000US-0232080.
PR	08-SEP-2000; 2000US-0232081.
PR	12-SEP-2000; 2000US-0231968.
PR	14-SEP-2000; 2000US-0232397.
PR	14-SEP-2000; 2000US-0232398.
PR	14-SEP-2000; 2000US-0232399.
PR	14-SEP-2000; 2000US-0232400.
PR	14-SEP-2000; 2000US-0232401.
PR	14-SEP-2000; 2000US-0233063.
PR	14-SEP-2000; 2000US-0233064.
PR	14-SEP-2000; 2000US-0233065.
PR	21-SEP-2000; 2000US-0234223.

PR	21-SEP-2000	2000US-02342474
PR	23-SEP-2000	2000US-02349978
PR	23-SEP-2000	2000US-02349987
PR	26-SEP-2000	2000US-02354834
PR	27-SEP-2000	2000US-02354834
PR	27-SEP-2000	2000US-02358362
PR	29-SEP-2000	2000US-02363527
PR	29-SEP-2000	2000US-02363567
PR	29-SEP-2000	2000US-02363568
PR	29-SEP-2000	2000US-02363569
PR	29-SEP-2000	2000US-02363702
PR	02-OCT-2000	2000US-02368070
PR	02-OCT-2000	2000US-02370337
PR	02-OCT-2000	2000US-02370338
PR	02-OCT-2000	2000US-02370339
PR	02-OCT-2000	2000US-02370400
PR	13-OCT-2000	2000US-02399353
PR	13-OCT-2000	2000US-02399357
PR	20-OCT-2000	2000US-02418008
PR	20-OCT-2000	2000US-02418009
PR	20-OCT-2000	2000US-02418221
PR	20-OCT-2000	2000US-02418226
PR	20-OCT-2000	2000US-02418865
PR	20-OCT-2000	2000US-02417877
PR	20-OCT-2000	2000US-02417877
PR	20-OCT-2000	2000US-02418008
PR	20-OCT-2000	2000US-02418221
PR	20-OCT-2000	2000US-02418226
PR	01-NOV-2000	2000US-02446177
PR	08-NOV-2000	2000US-02466474
PR	08-NOV-2000	2000US-02466475
PR	08-NOV-2000	2000US-02466476
PR	08-NOV-2000	2000US-02466477
PR	08-NOV-2000	2000US-02466478
PR	08-NOV-2000	2000US-02465528
PR	08-NOV-2000	2000US-02465528
PR	08-NOV-2000	2000US-02465524
PR	08-NOV-2000	2000US-02465524
PR	08-NOV-2000	2000US-02465525
PR	08-NOV-2000	2000US-02466110
PR	08-NOV-2000	2000US-02466111
PR	08-NOV-2000	2000US-02466133
PR	17-NOV-2000	2000US-02492077
PR	17-NOV-2000	2000US-02492078
PR	17-NOV-2000	2000US-02492080
PR	17-NOV-2000	2000US-02492090
PR	17-NOV-2000	2000US-02492100
PR	17-NOV-2000	2000US-02492111
PR	17-NOV-2000	2000US-02492112
PR	17-NOV-2000	2000US-02492123
PR	17-NOV-2000	2000US-02492123
PR	17-NOV-2000	2000US-02492125
PR	17-NOV-2000	2000US-02492125
PR	17-NOV-2000	2000US-02492126
PR	17-NOV-2000	2000US-02492127
PR	17-NOV-2000	2000US-02492128
PR	17-NOV-2000	2000US-02492244
PR	17-NOV-2000	2000US-02492245
PR	17-NOV-2000	2000US-02492245
PR	17-NOV-2000	2000US-02492246
PR	17-NOV-2000	2000US-02492265
PR	17-NOV-2000	2000US-02492297
PR	17-NOV-2000	2000US-02492599
PR	17-NOV-2000	2000US-02492599
PR	17-NOV-2000	2000US-02493000
PR	01-DEC-2000	2000US-02493000
PR	01-DEC-2000	2000US-02503191
PR	05-DEC-2000	2000US-02510330
PR	05-DEC-2000	2000US-02510388
PR	05-DEC-2000	2000US-02567179
PR	06-DEC-2000	2000US-02514719
PR	06-DEC-2000	2000US-02518656
PR	08-DEC-2000	2000US-02518658
PR	08-DEC-2000	2000US-02518668
PR	08-DEC-2000	2000US-02518689
PR	08-DEC-2000	2000US-02519899
PR	08-DEC-2000	2000US-02519899

PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Disclosure; SEQ ID NO 27042; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 47319 BP; 13708 A; 8453 C; 9720 G; 15438 T; 0 other;
 Query Match 67.9%; Score 22.4; DB 22; Length 47319;
 Best Local Similarity 81.2%; Pred. NO. 1.2e+02;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy 1 GCTGTGTTACTTTCTCTCTATTCAGCC 32
 Db 12933 GCTATGTCACTTACTTCTGATTCACCC 12902
 RESULT 5
 AAS11614
 ID AAS11614 standard; DNA; 236303 BP.
 XX
 AC AAS11614;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Human genomic DNA containing exons 2-17 of the CRIM1 gene.
 XX
 KM CRIM-1; Human; human chromosome 2p21-16.3; ophthalmological;
 KM neuroprotective; renal; osteopathic; dental; vulnaray; immunogen;
 KM antibody; gene therapy; neurodegenerative disease; eye disorder;
 KM cataract; bone morphogenic protein; BMG; renal disease; bone abnormality;
 KM tooth abnormality; wound; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FT 33104..33277 Location/Qualifiers
 FT
 FT /tag= a
 FT /number= 2
 FT 33278..77746
 FT /tag= b
 FT /number= 2
 FT 77747..77989
 FT /tag= c
 FT /number= 3
 FT 77990..79103
 FT intron

FT /tag= d
 FT /number= 3
 FT 79104..79224
 FT /tag= e
 FT /number= 4
 FT 79225..101022
 FT /tag= f
 FT /number= 4
 FT 101023..101144
 FT /tag= g
 FT /number= 5
 FT 101145..113377
 FT /tag= h
 FT /number= 5
 FT 113378..113560
 FT /tag= i
 FT /number= 6
 FT 113561..115985
 FT /tag= j
 FT /number= 6
 FT 115986..116183
 FT /tag= k
 FT /number= 7
 FT 116184..135707
 FT /tag= l
 FT /number= 7
 FT 135708..135836
 FT /tag= m
 FT /number= 8
 FT 135837..146471
 FT /tag= m
 FT /number= 8
 FT 146472..146628
 FT /tag= o
 FT /number= 9
 FT 146629..148761
 FT /tag= p
 FT /number= 9
 FT 148762..148883
 FT /tag= q
 FT /number= 10
 FT 148884..150044
 FT /tag= r
 FT /number= 10
 FT 150045..150254
 FT /tag= s
 FT /number= 11
 FT 150255..153815
 FT /tag= t
 FT /number= 11
 FT 153816..154031
 FT /tag= u
 FT /number= 12
 FT 154032..158580
 FT /tag= v
 FT /number= 12
 FT 158581..158802
 FT /tag= w
 FT /number= 13
 FT 158803..173982
 FT /tag= x
 FT /number= 13
 FT 173983..174177
 FT /tag= y
 FT /number= 14
 FT 174178..181006
 FT /tag= z
 FT /number= 14
 FT 181007..181129
 FT /tag= aa
 FT /number= 15
 FT 181130..183612
 FT /tag= ab
 FT intron

FT exon /number= 15
FT 183613..183800
FT /*cag= ac
FT /number= 16
FT intron 183801..185152
FT /*cag= ad
FT /number= 16
FT exon 185153..187765
FT /*cag= ae
FT /number= 17
XX MO200138519-A1.
XX
XX 31-MAY-2001.
XX
XX 24-NOV-2000; 2000MO-AU01435.
XX
XX 26-NOV-1999; 99AU-0004348.
XX
XX (UYQU) UNIV QUEENSLAND.
XX
XX Little M, Yamada T, Holmes G, Georgas K, Kolle G, Wilkinson L;
XX WPI; 2001-343951/36.
XX
XX Nucleic acids from human chromosome 2p21-16.3 and the encoded peptide,
XX useful for preventing, diagnosing and treating e.g. eye disease,
XX especially cataract formation -
XX
XX Claim 4; Fig 3; 169pp; English.
XX
XX The invention relates to nucleic acids from human chromosome 2p21-16.3
XX and the encoded peptide (and mouse and chicken orthologues) that
XX comprises a PGCeCIP group, an insulin-like growth factor binding protein
XX (IGFBP)-like domain, cysteine-rich domains, an RGD (undefined) group
XX and a transmembrane domain. The protein, e.g. CRIM1, interacts with
XX peptides of the transforming growth factor superfamily. A composition
XX comprising an expression construct comprising the nucleic acids of the
XX invention or a mimetic which antagonises or mimics an activity of a CRIM1
XX polypeptide may be used in a method for modulating the biological
XX activity of a polypeptide of the bone morphogenic protein (BMP) family.
XX In this way they may be used to prevent or treat an eye disease,
XX especially cataract formation. They may also be used to treat
XX neurodegenerative diseases, renal and kidney disease, bone and tooth
XX abnormalities, wounds and skin damage, e.g. by use of the nucleic acid in
XX gene therapy by using antibodies directed against CRIM1 polypeptides.
XX The present sequence is a Human genomic DNA containing exons 2-17 of
XX the CRIM1 gene.
XX
XX Sequence 236303 BP; 70199 A; 46544 C; 47996 G; 71563 T; 1 other;
XX
XX Query Match 67.9%; Score 22.4; DB 22; Length 236303;
XX Best Local Similarity 81.2%; Pred. No. 1.4e+02;
XX Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
XX
XX 1 GCTGTGTACTTTCTTCTTCAATCCAGCC 32
XX |||||
Db 214710 GCTGATGCTACTTACTTCTTGTATTCACAC 214741
XX
XX RESULT 6
XX AAC43567
XX ID AAC43567 standard; DNA; 458 BP.
XX
XX AAC43567;
XX
XX 17-OCT-2000 (first entry)
XX
XX Zea mays DNA fragment SEQ ID NO: 39699.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic;
XX pathway; promoter; termination sequence; corn; ss.

XX OS Zea mays subsp. mays.
XX PN EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142350.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-013624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147058.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 66.7%; Score 22; DB 21; Length 458;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GCTGTTGTAATTTCTTCTTATTCAG 30
Db 114 GCTTTGTAATCTTCTTCTTCTTCCAG 143

RESULT 7
ID ABA18638/c
ID ABA18638 standard; DNA; 13079 BP.
XX
AC ABA18638;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 10969.
XX
KW Human; nootropic; neuroprotective; cytosolic; dermatological; virologic;
KW immunosuppressive; antilinfimatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisticking; antianaemic; antiarthritic; cancer;
KW antineumatic; hepatocrotic; cerebroprotective; antinflamatory;
KW antiallergic; antidiabetic; antilicer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; de.
XX
OS Homo sapiens.
XX

PN WC200159063-A2.
 XX 16-AUG-2001.
 PD
 XX
 PF 17-JAN-2001; 2001WO-US01334.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217949.
 PR 26-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244617.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251997.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SW;
 PI

XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides.
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX Disclosure; SEQ ID NO 10969; 1701bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABAI21534) and proteins
CC (ABAI1678-ABAI18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 13079 BP; 4243 A; 2205 C; 2671 G; 3960 T; 0 other;
Query Match 66.7%; Score 22; DB 22; Length 13079;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GCTGTGTTACTTTCTCTCTATTCAG 30
DB 11911 GGTGTGTACATTTCTTTCTATCAAG 11882
RESULT 8
ABAI18639/C
ID ABAI18639 standard; DNA; 25967 BP.
XX
AC ABAI18639;
XX
XX 23-JAN-2002 (first entry)
DT
XX
XX Human nervous system related polynucleotide SEQ ID NO 10970.
DE
XX
XX Human; goitrotic; neuroprotective; cytostatic; dermatological; virocidic;
KM immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulterary;
KM antiparkinsonian; antichilling; antianemic; antiarthritic; cancer;
KM antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KM antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KM antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KM neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216687.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 25-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 14-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0246417.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX
PS Disclosure; SEQ ID NO 10970; 1701bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful

CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 25967 BP; 8245 A; 4194 C; 4708 G; 8820 T; 0 other;
Query Match 66.7%; Score 22; DB 22; Length 25967;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 11919 GGTGTGTACATTCCTTCTTATCAG 11890
Gy 1 GCTGTGTACTTCTTCTTCTTATTCAG 30
ID ABA75350 standard; DNA; 144 BP.
AC ABA75350;
XX
XX 01-FEB-2002 (first entry)
DT
XX
XX Human foetal liver single exon nucleic acid probe #23655.
DE
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
KM
XX
XX Homo sapiens.
OS
XX
XX NO200157277-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX
XX 30-JAN-2001; 2001MO-US00669.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human foetal liver -
XX
XX
XX Claim 4; SEQ ID NO 23655; 639bp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

```

SQ      Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
QY      Query Match          65.5%; Score 21.6; DB 22; Length 144;
        Best Local Similarity 85.7%; Pred. No. 1.6e+02;
        Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY      3 TGTGTACTTCTCTCTATTCAG 30
        | ||||| ||||| ||||| |||||
DB      29 TTTCGTTCTTCTCTCTCTCTAG 56

RESULT 10
ABA33992
ID      ABA33992 standard; DNA; 144 BP.
XX
XX      ABA33992;
XX
XX      23-JAN-2002 (first entry)
DE      Probe #18458 for gene expression analysis in human heart cell sample.
XX
XX      Human; gene expression; heart; microarray; vascular system; probe;
KW      cardiovascular disease; hypertension; cardiac arrhythmia;
XX      congenital heart disease; ss.
XX
XX      Homo sapiens.
OS
PN      WO200157274-A2.
XX
XX      09-AUG-2001.
PD
PF      30-JAN-2001; 2001WO-US00666.
XX
XX      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
XX      (MOLE-) MOLECULAR DYNAMICS INC.
PA
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX      WPI; 2001-488899/53.
DR
XX      Single exon nucleic acid probes for analyzing gene expression in human
PT      hearts -
PT
PS      Claim 4; SEQ ID No 18458; 530bp; English.
XX
XX
CC      The present invention relates to single exon nucleic acid probes for
CC      measuring human gene expression in a sample derived from human heart. The
CC      present sequence is one such probe. The probes may be used for
CC      predicting, measuring and displaying gene expression in samples derived
CC      from the human heart via microarrays. By measuring gene expression, the
CC      probes are useful for predicting, diagnosing, grading, staging,
CC      monitoring and prognosing diseases of the human heart and vascular system
CC      e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC      congenital heart disease.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
QY      Query Match          65.5%; Score 21.6; DB 22; Length 144;
        Best Local Similarity 85.7%; Pred. No. 1.6e+02;
        Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY      3 TGTGTACTTCTCTCTATTCAG 30

```

```

Db      29 TTTTCTTCTTCTCTCTCTCTCTCTAG 56

RESULT 11
ID      AAK23906
XX      AAK23906 standard; DNA; 144 BP.
AC      AAK23906;
XX
DT      05-NOV-2001 (first entry)
XX
DE      Human brain expressed single exon probe SEQ ID NO: 23897.
XX
KW      Human; brain expressed exon; gene expression analysis; probe;
KW      microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW      epilepsy; cancer; ss.
XX
OS      Homo sapiens.
XX
MO      MO200157275-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US00667.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-483446/52.
XX
PT      Single exon nucleic acid probes for analyzing gene expression in human
PT      brains -
XX
PS      Example 4; SEQ ID NO: 23897; 650bp + Sequence Listing; English.
XX
CC      The present invention provides a number of single exon nucleic acid
CC      probes which are derived from genomic sequences expressed in the human
CC      brain. They can be used to measure gene expression in brain cell samples,
CC      which may enable the diagnosis and improved treatment of nervous system
CC      diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC      epilepsy and cancers. The present sequence is one of the probes of the
CC      invention.
XX
SQ      Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
XX
Query Match      65.5%; Score 21.6; DB 22; Length 144;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy      3 TGTGTTACTTTCTCTCTATTCAG 30
Db      29 TTTTGTTCTTCTCTCTCTCTCTAG 56

RESULT 12
ID      AAK49980
XX      AAK49980 standard; DNA; 144 BP.
AC      AAK49980;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human bone marrow expressed single exon probe SEQ ID NO: 24537.
XX

```

XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX Homo sapiens.
 OS
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 24537; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 SQ Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
 XX
 Query Match 65.5%; Score 21.6; DB 22; Length 144;
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 3 TGTGTTACTTTCTTCTTCTTCTATTCGAG 30
 Db 29 TTTGTTCTTTCTTCTTCTTCTTCTTAG 56
 XX
 RESULT 13
 ID AAI27077 standard; DNA; 144 BP.
 AC AAI27077;
 XX
 DT 12-OCT-2001 (first entry)
 XX
 DE Probe #17010 for gene expression analysis in human cervical cell sample.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.
 XX
 OS Homo sapiens.
 OS
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00670.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR

PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 25; SEQ ID No 17010; 487bp; English.
 XX
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENPs). The present sequence is one such probe. The SENPs are derived
 CC from human Hela cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 144 BP; 27 A; 33 C; 12 G; 72 T; 0 other;
 XX
 Query Match 65.5%; Score 21.6; DB 22; Length 144;
 Best Local Similarity 85.7%; Pred. No. 1.6e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 3 TGTGTTACTTTCTTCTTCTTCTATTCGAG 30
 Db 29 TTTGTTCTTTCTTCTTCTTCTTCTTAG 56
 XX
 RESULT 14
 ID ABA62842 standard; DNA; 568 BP.
 AC ABA62842;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #11147.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 OS
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 11147; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 568 BP; 196 A; 95 C; 69 G; 208 T; 0 other;

Query Match 65.5%; Score 21.6; DB 22; Length 568;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TGTGTACTTTCTTCTTCTATTCAG 30
Db 97 TTTTGTCTTTCTTCTTCTTCTAG 124

RESULT 15
ABA30124
ID ABA30124 standard; DNA; 568 BP.
XX
AC ABA30124;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #8590 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PR 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 1; SEQ ID No 8590; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system

CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 568 BP; 196 A; 95 C; 69 G; 208 T; 0 other;

Query Match 65.5%; Score 21.6; DB 22; Length 568;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TGTGTACTTTCTTCTTCTATTCAG 30
Db 97 TTTTGTCTTTCTTCTTCTTCTAG 124

Search completed: May 11, 2003, 03:06:50
Job time : 108.18 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 / Search time 13.231 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-16

Perfect score: 33

Sequence: 1 gctgtgtctctctctctctctcagcaca 33

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

- 1: Issued_Patents_NA.*
- 2: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20.4	61.8	26173	4	US-09-453-702B-69 Sequence 69, App1
C 2	19.8	60.0	4145	4	US-09-302-620B-82 Sequence 82, App1
C 3	19.8	60.0	4206	4	US-09-302-620B-81 Sequence 81, App1
C 4	19.6	59.4	2623	2	US-08-973-675-1 Sequence 59, App1
C 5	19.6	59.4	14855	2	US-08-687-080-59 Sequence 15, App1
C 6	19	57.6	1314	1	US-07-662-005A-15 Sequence 2, App1
C 7	19	57.6	1351	4	US-09-002-114-2 Sequence 7, App1
C 8	18.8	57.0	594	4	US-09-687-477-7 Sequence 11, App1
C 9	18.8	57.0	594	4	US-09-687-477-13 Sequence 15, App1
C 10	18.8	57.0	594	4	US-09-687-477-15 Sequence 6, App1
C 11	18.8	57.0	594	4	US-09-687-477-13 Sequence 11, App1
C 12	18.8	57.0	594	4	US-09-687-477-13 Sequence 15, App1
C 13	18.6	56.4	45346	4	US-09-146-053-6 Sequence 1709, App1
C 14	18.6	56.4	1191	4	US-09-134-001C-1709 Sequence 3, App1
C 15	18.6	56.4	2196	3	US-08-472-217-3 Sequence 3, App1
C 16	18.6	56.4	2196	3	US-08-472-217-3 Sequence 3, App1
C 17	18.6	56.4	7350	1	US-07-865-662F-14 Sequence 14, App1
C 18	18.6	56.4	7350	1	US-08-374-219B-16 Sequence 16, App1
C 19	18.6	56.4	168575	4	US-09-426-280-1 Sequence 1, App1
C 20	18.4	55.8	1284	4	US-09-134-001C-212 Sequence 212, App1
C 21	18.4	55.8	2543	4	US-09-307-143-5 Sequence 5, App1
C 22	18.4	55.8	3387	1	US-08-468-557-1 Sequence 1, App1
C 23	18.2	55.2	17327	1	US-07-906-871-15 Sequence 15, App1
C 24	18.2	55.2	223	4	US-08-205-697A-14 Sequence 14, App1
C 25	18.2	55.2	223	4	US-08-205-697A-14 Sequence 14, App1
C 26	18.2	55.2	301	4	PCT-US95-02576-14 Sequence 14, App1
C 27	18.2	55.2	301	4	US-09-605-785-234 Sequence 234, App1
					US-09-439-313-234 Sequence 234, App1

28	18.2	55.2	301	4	US-09-352-616A-234 Sequence 234, App
29	18.2	55.2	301	4	US-09-232-149A-234 Sequence 234, App
30	18.2	55.2	503	4	US-09-280-116-203 Sequence 203, App
31	18.2	55.2	589	4	US-09-328-111-546 Sequence 546, App
32	18.2	55.2	659	2	US-08-718-538-4 Sequence 4, App1
33	18.2	55.2	890	2	US-08-319-576-1 Sequence 1, App1
34	18.2	55.2	1005	2	US-08-701-935-4 Sequence 4, App1
35	18.2	55.2	1005	3	US-09-134-591-4 Sequence 4, App1
36	18.2	55.2	1261	4	US-08-205-697A-12 Sequence 12, App1
37	18.2	55.2	1261	4	US-08-702-525-12 Sequence 12, App1
38	18.2	55.2	1261	5	PCT-US95-02576-12 Sequence 12, App1
39	18.2	55.2	1509	4	US-09-149-476-179 Sequence 179, App
40	18.2	55.2	2896	1	US-08-441-430-31 Sequence 31, App1
41	18.2	55.2	2995	1	US-08-441-430-32 Sequence 32, App1
42	18.2	55.2	3387	1	US-08-064-271-11 Sequence 11, App1
43	18.2	55.2	3387	2	US-08-627-554C-29 Sequence 29, App1
44	18.2	55.2	3387	3	US-08-930-589A-19 Sequence 19, App1
45	18.2	55.2	3387	4	US-09-599-781-19 Sequence 19, App1

ALIGNMENTS

RESULT 1
US-09-453-702B-69/c
Sequence 69, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blattner, Frederick R.
Burland, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453, 702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 26173
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 69:
US-09-453-702B-69
Query Match 61.8%; Score 20.4; DB 4; Length 26173;
Best Local Similarity 80.0%; Pred. No. 58;

Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 4 GTGTACTTTCTCTCTATTCAGCA 33

Db 8207 GTGTACTTTCTCTCTATTCAGCA 8178

RESULT 2

US-09-302-6208-82/c

Sequence 82, Application US/093026208

Patent No. 6331420

GENERAL INFORMATION:

APPLICANT: Wilson, C. Ron

APPLICANT: Craft, David L.

APPLICANT: Eirich, Dudley

APPLICANT: Eshoo, Mark

APPLICANT: Madduri, Krishna M.

APPLICANT: Cornett, Cathy A.

APPLICANT: Brenner, Alfred A.

APPLICANT: Tang, Maria

APPLICANT: Loper, John C.

TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450

TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA

TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS

FILE REFERENCE: 1010-16, seq

CURRENT APPLICATION NUMBER: US/09/302,6208

NUMBER OF SEQ ID NOS: 109

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 82

LENGTH: 4145

TYPE: DNA

ORGANISM: Candida tropicalis

US-09-302-6208-82

Query Match 60.0%; Score 19.8; DB 4; Length 4145;

Best Local Similarity 77.4%; Pred. No. 73;

Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 3 TGTGTACTTTCTCTCTATTCAGCA 33

Db 1207 TGTGTACTTTCTCTCTATTCAGCA 1177

RESULT 3

US-09-302-6208-81/c

Sequence 81, Application US/093026208

Patent No. 6331420

GENERAL INFORMATION:

APPLICANT: Wilson, C. Ron

APPLICANT: Craft, David L.

APPLICANT: Eirich, Dudley

APPLICANT: Eshoo, Mark

APPLICANT: Madduri, Krishna M.

APPLICANT: Cornett, Cathy A.

APPLICANT: Brenner, Alfred A.

APPLICANT: Tang, Maria

APPLICANT: Loper, John C.

TITLE OF INVENTION: CYTOCHROME P450 MONOOXYGENASE AND NADPH CYTOCHROME P450

TITLE OF INVENTION: OXIDOREDUCTASE GENES AND PROTEINS RELATED TO THE OMEGA

TITLE OF INVENTION: HYDROXYLASE COMPLEX OF CANDIDA TROPICALIS AND METHODS

FILE REFERENCE: 1010-16, seq

CURRENT APPLICATION NUMBER: US/09/302,6208

NUMBER OF SEQ ID NOS: 109

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 81

LENGTH: 4206

TYPE: DNA

ORGANISM: Candida tropicalis

US-09-302-6208-81

Query Match 60.0%; Score 19.8; DB 4; Length 4206;

Best Local Similarity 77.4%; Pred. No. 73;

Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 3 TGTGTACTTTCTCTCTATTCAGCA 33

Db 1180 TGTGTACTTTCTCTCTATTCAGCA 1150

RESULT 4

US-08-973-675-1/c

Sequence 1, Application US/08973675

Patent No. 5985283

GENERAL INFORMATION:

APPLICANT: HATEBOER, GUNS

APPLICANT: BERNARDS, RENE

TITLE OF INVENTION: ADENOVIRUS E1A-ASSOCIATED PROTEIN BS69,

TITLE OF INVENTION: INHIBITOR OF E1A-TRANSACTIVATION

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/973,675

FILING DATE: 12-DEC-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: WILSON, MARY J.

REGISTRATION NUMBER: 32,955

REFERENCE/DOCKET NUMBER: 620-31

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2623 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

FEATURE:

NAME/KEY: CDS

LOCATION: 150..1835

US-08-973-675-1

Query Match 59.4%; Score 19.6; DB 2; Length 2623;

Best Local Similarity 84.4%; Pred. No. 81;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 5 TTGTACTTTCTCTCTATTCAG 30

Db 1562 TTGTACTTTCTCTCTATTCAG 1537

RESULT 5

US-08-687-080-59

Sequence 59, Application US/08687080

Patent No. 5965427

GENERAL INFORMATION:

APPLICANT: Gregory Dolganov

TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof

NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 14855 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 5' END OF INTRON 2 OF RAD50 GENOMIC
INDIVIDUAL ISOLATE: SEQUENCE
US-08-687-080-59

Query Match 59.4%; Score 19.6; DB 2; Length 14855;
Best Local Similarity 84.6%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

CY 8 TTACTTTCTTCTTATTCAGCCA 33
DB 5760 TTACTTTCTTCTTCTCAGCCA 5785

RESULT 6
US-07-662-005A-15
Sequence 15, Application US/07662005A
Patent No. 5246838
GENERAL INFORMATION:
APPLICANT: Van Dijk, Jan M.
APPLICANT: Smith, Hilde E.
APPLICANT: Bron, Sierd
APPLICANT: Quax, Wilhelmus J.
TITLE OF INVENTION: IMPROVED PROCESSING OF PROTEINS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/662,005A
FILING DATE: 19910228
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Muraahige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 246152002300
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600
TELEFAX: 415-494-0792
TELEX: 34-0154
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1314 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-662-005A-15

Query Match 57.6%; Score 19; DB 1; Length 1314;
Best Local Similarity 81.5%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

CY 4 GTTGTTACTTTCTTCTTATTCAG 30
DB 554 GTAGTTACTTCTTCTTCTTCTTACAG 580

RESULT 7
US-09-002-114-2/c
Sequence 2, Application US/09002114
Patent No. 6274720
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purni
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN PREPRONEUROTENSIN/NEUROMEDIN N
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/002,114
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0450 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1351 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-002-114-2

Query Match 57.6%; Score 19; DB 4; Length 1351;
Best Local Similarity 81.5%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTCTCTCTATTC 28

Db 1306 CTCTTTTACTTTCTTTCTTTTC 1280

RESULT 8

US-09-687-477-7
; Sequence 7, Application US/09687477
; Patent No. 6463883
; GENERAL INFORMATION:
; APPLICANT: Aquabio Product Sciences, LLC
; APPLICANT: Harris, H. William, Jr.
; APPLICANT: Russell, David R.
; APPLICANT: Neering, Jacqueline
; APPLICANT: Betka, Marlies
; TITLE OF INVENTION: Methods for Raising Pre-Adult Anadromous
; TITLE OF INVENTION: Fish
; FILE REFERENCE: 2213.1004-000
; CURRENT APPLICATION NUMBER: US/09/687,477
; CURRENT FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Chum Salmon
US-09-687-477-7

Query Match 57.0%; Score 18.8; DB 4; Length 594;
Best Local Similarity 76.7%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTCTCTCTATTCAGC 31

Db 119 CTGTTCTCACTTATCTGCTGTTTCCAGC 148

RESULT 9

US-09-687-477-11
; Sequence 11, Application US/09687477
; Patent No. 6463883
; GENERAL INFORMATION:
; APPLICANT: Aquabio Product Sciences, LLC
; APPLICANT: Harris, H. William, Jr.
; APPLICANT: Russell, David R.
; APPLICANT: Neering, Jacqueline
; APPLICANT: Betka, Marlies
; TITLE OF INVENTION: Methods for Raising Pre-Adult Anadromous
; TITLE OF INVENTION: Fish
; FILE REFERENCE: 2213.1004-000
; CURRENT APPLICATION NUMBER: US/09/687,477
; CURRENT FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 594
; TYPE: DNA
; ORGANISM: King Salmon
US-09-687-477-11

Query Match 57.0%; Score 18.8; DB 4; Length 594;
Best Local Similarity 76.7%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTCTCTCTATTCAGC 31

Db 119 CTGTTCTCACTTATCTGCTGTTTCCAGC 148

RESULT 10
US-09-687-477-13
; Sequence 13, Application US/09687477
; Patent No. 6463883
; GENERAL INFORMATION:
; APPLICANT: Aquabio Product Sciences, LLC
; APPLICANT: Harris, H. William, Jr.
; APPLICANT: Russell, David R.
; APPLICANT: Neering, Jacqueline
; APPLICANT: Betka, Marlies
; TITLE OF INVENTION: Methods for Raising Pre-Adult Anadromous
; TITLE OF INVENTION: Fish
; FILE REFERENCE: 2213.1004-000
; CURRENT APPLICATION NUMBER: US/09/687,477
; CURRENT FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Pink Salmon
US-09-687-477-13

Query Match 57.0%; Score 18.8; DB 4; Length 594;
Best Local Similarity 76.7%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTCTCTCTATTCAGC 31

Db 119 CTGTTCTCACTTATCTGCTGTTTCCAGC 148

RESULT 11

US-09-687-477-15
; Sequence 15, Application US/09687477
; Patent No. 6463883
; GENERAL INFORMATION:
; APPLICANT: Aquabio Product Sciences, LLC
; APPLICANT: Harris, H. William, Jr.
; APPLICANT: Russell, David R.
; APPLICANT: Neering, Jacqueline
; APPLICANT: Betka, Marlies
; TITLE OF INVENTION: Methods for Raising Pre-Adult Anadromous
; TITLE OF INVENTION: Fish
; FILE REFERENCE: 2213.1004-000
; CURRENT APPLICATION NUMBER: US/09/687,477
; CURRENT FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Sockeye Salmon
US-09-687-477-15

Query Match 57.0%; Score 18.8; DB 4; Length 594;
Best Local Similarity 76.7%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 CTGTTGTTACTTTCTCTCTATTCAGC 31

Db 119 CTGTTCTCACTTATCTGCTGTTTCCAGC 148

RESULT 12

US-09-146-053-6/c
; Sequence 6, Application US/09146053A
; Patent No. 6399349
; GENERAL INFORMATION:
; APPLICANT: Ryan, James W.
; APPLICANT: Sprinkle, Terry Joe Curtis
; APPLICANT: Venema, Richard C.

;; TITLE OF INVENTION: Human Amino-peptidase P Gene
;; FILE REFERENCE: MCG103
;; CURRENT APPLICATION NUMBER: US/09/146,053A
;; CURRENT FILING DATE: 1998-09-02
;; EARLIER APPLICATION NUMBER: 60/057,854
;; EARLIER FILING DATE: 1997-09-02
;; NUMBER OF SEQ ID NOS: 7
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO: 6
;; LENGTH: 45546
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-146-053-6

Query Match 57.0%; Score 18.8; DB 4; Length 45546;
Best Local Similarity 76.7%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TGTGTACTTCTTCTCTTCTATTCAGCC 32
DB 10468 TGACTTAGTTTCATCTCTATCCCGCC 10439

RESULT 13
US-09-134-001C-1709/c
; Sequence 1709, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO: 1709
; LENGTH: 1191
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1709

Query Match 56.4%; Score 18.6; DB 4; Length 1191;
Best Local Similarity 84.0%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTGTGTACTTCTTCTCTCTATT 26
DB 837 CTGTCTTCTTATCTTCTCTATT 813

RESULT 14
US-08-472-217-3
; Sequence 3, Application US/08472217
; Patent No. 5726058
; GENERAL INFORMATION:
; APPLICANT: Alanen-Kurki, Leena
; APPLICANT: Auvunen, Petri
; APPLICANT: Jaakkola, Panu
; APPLICANT: Jaakkonen, Markku
; APPLICANT: Lepp, Sirpa
; APPLICANT: Maki, Markku
; APPLICANT: Viininen, Tapani
; APPLICANT: Wrti, Anni
; TITLE OF INVENTION: Syndecan Stimulation Of Cellular
; TITLE OF INVENTION: Differentiation
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600

;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20005
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/472,217
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/206,186
;; FILING DATE: 07-MAR-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/988,427
;; FILING DATE: 01-DEC-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Ciabala, Michele A.
;; REGISTRATION NUMBER: 33,851
;; REFERENCE/DOCKET NUMBER: 1102.0050003
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2196 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: both
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
US-08-472-217-3

Query Match 56.4%; Score 18.6; DB 1; Length 2196;
Best Local Similarity 72.7%; Pred. No. 1.8e+02;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GGTGTGTACTTCTTCTCTTCTATTCAGCCA 33
DB 224 GTGTCTCTCTTCTCTCTCTCTGTTCGCA 256

RESULT 15
US-08-760-534A-3
; Sequence 3, Application US/08760534A
; Patent No. 6017727
; GENERAL INFORMATION:
; APPLICANT: JALKANEN, MARKKU
; APPLICANT: JAAKKOLA, PANU
; APPLICANT: VIININEN, TAPANI
; TITLE OF INVENTION: SYNDECAN ENHANCER ELEMENT AND SYNDECAN
; TITLE OF INVENTION: STIMULATION OF CELLULAR DIFFERENTIATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/760,534A
; FILING DATE: 02-DEC-1996
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/206,186
 ; FILING DATE: 07-MAR-1994
 ; PRIOR APPLICATION DATA: PCT/FI93/00514
 ; APPLICATION NUMBER: PCT/FI93/00514
 ; FILING DATE: 01-DEC-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: CIMBALA, MICHELE A.
 ; REGISTRATION NUMBER: 33,851
 ; REFERENCE/DOCKET NUMBER: 1708.0050004/MAC
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 371-2600
 ; TELEFAX: (202) 371-2540
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2196 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-760-534A-3

Query Match 56.4%; Score 18.6; DB 3; Length 2196;
 Best Local Similarity 72.7%; Pred. No. 1.8e+02;
 Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GCTGTTGTTACTTTCTCTCTCTATTCAGCA 33
 Db 224 GTTGTGCTTCTTTCTCTCTCTCTGTTCTGCA 256

Search completed: May 11, 2003, 03:08:01
 Job time : 27.2331 secs

APPLICANT: LITTLE, Melissa

```

RESULT 2
US-10-152-724A-24
: Sequence 24, Application US/10152724A
: Publication No. US20030082714A1
: GENERAL INFORMATION:
: APPLICANT: LITTLE, Melissa
: APPLICANT: HOLMES, Gregory

```

APPLICANT: KOLLE, Gabriel
APPLICANT: YAMADA, Toshiya
APPLICANT: GEORGAS, Kyli
APPLICANT: WILKINSON, Lorine
TITLE OF INVENTION: No. US20030082714A1 Nucleic Acid and Polypeptide
FILE REFERENCE: P22378
CURRENT APPLICATION NUMBER: US/10/152,724A
CURRENT FILING DATE: 2002-05-23
PRIOR APPLICATION NUMBER: Australian App No. US20030082714A1PQ 4348
PRIOR FILING DATE: 1999-11-26
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.1
SEQ ID NO 24
LENGTH: 36303
TYPE: DNA
ORGANISM: Homo sapiens
US-10-152-724A-24

Query Match 67.9%; Score 22.4; DB 9; Length 36303;
Best Local Similarity 81.2%; Pred. No. 82;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 GCTGTGTTACTTTCTTCTCTATTCAGCC 32
Db 14710 GCTGATGTCACCTTACTTGTGATTCACACC 14741

RESULT 3
US-09-864-761-25312
Sequence 25312, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemlica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 25312
LENGTH: 144
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL121953.10
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EST HUMAN HIT: BE786184.1, EVALUATE 1.00e-20
OTHER INFORMATION: NT HIT: g17706668, EVALUATE 7.00e-21
US-09-864-761-25312

Query Match 65.5%; Score 21.6; DB 10; Length 144;
Best Local Similarity 85.7%; Pred. No. 64;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TGTGTACTTTCTTCTCTATTCAG 30
Db 29 TTTGTGTTCTTTCTTCTCTCTCTAG 56

RESULT 4
US-09-796-692-6086/c
Sequence 6086, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannon, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
FILE REFERENCE: 2077,001200
CURRENT APPLICATION NUMBER: US/09/796,692
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: 60/223,378
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 9597
SOFTWARE: FastsEQ for Windows Version 3.0
SEQ ID NO 6086
LENGTH: 310
TYPE: DNA


```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (184)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (201)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (213)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (217)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (260)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (281)
OTHER INFORMATION: n=A,T,C or G
NAME/KEY: unsure
LOCATION: (308)
OTHER INFORMATION: n=A,T,C or G
US-09-796-692-6086
```

```
Query Match      65.5%; Score 21.6; DB 9; Length 310;
Best Local Similarity 85.7%; Pred. No. 72;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 3 TGTGTACTTTCTTCTTCTATTCAG 30
Db 70 TTTGTCTTCTTCTTCTTCTCTAG 43
```

```
RESULT 5
US-10-040-862-6086/c
Sequence 6086, Application US/10040862
Publication No. US20030078396A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Retter, Marc
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
FILE REFERENCE: 014058-013520US
CURRENT APPLICATION NUMBER: US/10/040,862
PRIOR FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: US 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: US 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: US 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: US 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: US 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: US 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US 60/223,378
PRIOR FILING DATE: 2000-08-07
```

```
PRIOR APPLICATION NUMBER: US 09/796,692
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10467
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6086
LENGTH: 310
TYPE: DNA
```

```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (184)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (201)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (213)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (217)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (260)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (281)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (308)
OTHER INFORMATION: n=A,T,C or G
US-10-040-862-6086
```

```
Query Match      65.5%; Score 21.6; DB 9; Length 310;
Best Local Similarity 85.7%; Pred. No. 72;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
OY 3 TGTGTACTTTCTTCTTCTATTCAG 30
Db 70 TTTGTCTTCTTCTTCTTCTCTAG 43
```

```
RESULT 6
US-09-796-692-4667/c
Sequence 4667, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Corixa Corporation
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
FILE REFERENCE: 2077,001200
CURRENT APPLICATION NUMBER: US/09/796,692
PRIOR FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/202,084
```

```
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4667
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4667

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TGTGTACTTTCTTCTTCTTCTTCCAG 30
    ||||| ||||| ||||| ||||| |||||
Db 70 TTTGTTCTTTCTTCTTCTTCTTCTAG 43

RESULT 7
US-09-796-692-5376/c
; Sequence 5376, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077, 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5376
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
NAME/KEY: unsure
```

```
; LOCATION: (165)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-5376

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TGTGTACTTTCTTCTTCTTCTTCCAG 30
    ||||| ||||| ||||| ||||| |||||
Db 70 TTTGTTCTTTCTTCTTCTTCTTCTAG 43

RESULT 8
US-09-796-692-7106/c
; Sequence 7106, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077, 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7106
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-7106

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TGTGTACTTTCTTCTTCTTCTTCCAG 30
    ||||| ||||| ||||| ||||| |||||
Db 70 TTTGTTCTTTCTTCTTCTTCTTCTAG 43

RESULT 9
US-10-040-862-4667/c
; Sequence 4667, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
```

```
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane
APPLICANT: Retter, Marc
APPLICANT: Cortix Corporation
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
FILE REFERENCE: 014058-013520US
CURRENT APPLICATION NUMBER: US/10/040,862
CURRENT FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: US 60/186,126
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: US 60/190,479
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: US 60/200,545
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: US 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: US 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: US 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: US 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US 60/223,378
PRIOR FILING DATE: 2000-08-07
PRIOR APPLICATION NUMBER: US 09/796,692
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10467
SOFTWARE: FastSeq for Windows Version:3.0
SEQ ID NO 4667
LENGTH: 408
TYPE: DNA
ORGANISM: Homo sapiens
US-10-040-862-4667

Query Match          65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTTACTTTCTTCTTCTATTCAG 30
Db 70 TTTGTTCTTTCTTCTTCTTCTTAG 43

RESULT 10
US-10-040-862-5376/c
; Sequence 5376, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Cortix Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
```

```
PRIOR APPLICATION NUMBER: US 60/200,303
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,779
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: US 60/200,999
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: US 60/202,084
PRIOR FILING DATE: 2000-05-04
PRIOR APPLICATION NUMBER: US 60/206,201
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: US 60/218,950
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 60/222,903
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/223,416
PRIOR FILING DATE: 2000-08-04
PRIOR APPLICATION NUMBER: US 60/223,378
PRIOR FILING DATE: 2000-08-07
PRIOR APPLICATION NUMBER: US 09/796,692
PRIOR FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 10467
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5376
LENGTH: 408
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: (165)
OTHER INFORMATION: n=A,T,C or G
US-10-040-862-5376
```

```
Query Match          65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTTACTTTCTTCTTCTATTCAG 30
Db 70 TTTGTTCTTTCTTCTTCTTCTTAG 43
```

```
RESULT 11
US-10-040-862-7106/c
; Sequence 7106, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Cortix Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
```

```

; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: PaedSeq for Windows Version 3.0
; SEQ ID NO 7106
; LENGTH: 408
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-7106

Query Match      65.5%; Score 21.6; DB 9; Length 408;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 TGTGTACTTTCTTCTTCTATCCAG 30
DB      70 TTGTGTTCTTTCTTCTTCTCTCTAG 43

RESULT 12
US-09-783-590-151/C
; Sequence 151, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven W.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 151
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (284)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (324)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (344)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (369)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (402)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature

```

```

; LOCATION: (407)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (413)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (419)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (428)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (436)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (439)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (449)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (450)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (457)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (461)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (482)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-151

Query Match      65.5%; Score 21.6; DB 10; Length 499;
Best Local Similarity 85.7%; Pred. No. 78;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 TGTGTACTTTCTTCTTCTATCCAG 30
DB      85 TTGTGTTCTTTCTTCTTCTCTCTTCTG 58

RESULT 13
US-09-864-761-8590
; Sequence 8590, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30

```

PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 8590
LENGTH: 568
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL121953.10
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
US-09-864-761-8590

Query Match 65.5%; Score 21.6; DB 10; Length 568;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTACTTTCTTCTTCTATTCAG 30
DB 97 TTTGTTCCTTCTTCTTCTCTAG 124

RESULT 14
US-09-822-846-293/C
Sequence 293, Application US/09822846
Publication No. US20030027139A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallie, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steininger II, Robert J.
APPLICANT: Bowman, Michael R.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Pechtel, Kim
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakkar
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6400
CURRENT APPLICATION NUMBER: US/09/822,846
CURRENT FILING DATE: 2001-03-29

PRIOR APPLICATION NUMBER: 60/195,605
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 629
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 293
LENGTH: 1057
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-846-293

Query Match 65.5%; Score 21.6; DB 9; Length 1057;
Best Local Similarity 85.7%; Pred. No. 88;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGTGTACTTTCTTCTTCTATTCAG 30
DB 549 TTTGTTCCTTCTTCTTCTCTAG 522

RESULT 15
US-09-938-842A-5185
Sequence 5185, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 5185
LENGTH: 356
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-5185

Query Match 64.8%; Score 21.4; DB 9; Length 356;
Best Local Similarity 80.6%; Pred. No. 87;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 TGTGTACTTTCTTCTTCTATTCAGCA 33
DB 277 TGTGTTCCTTCTTCTTCTATTCAGCA 307

Search completed: May 11, 2003, 06:28:01
Job time : 36.6692 secs

Mon May 12 14:15:08 2003

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 328.336 Seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-17

Sequence: 1 attccacattcattccgcccttcattg 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_hgt: *
3: gb_in: *
4: gb_cm: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hgt_hum: *
31: em_hgt_inv: *
32: em_hgt_other: *
33: em_hgt_mus: *
34: em_hgt_pln: *
35: em_hgt_rtd: *
36: em_hgt_mam: *
37: em_hgt_vtc: *
38: em_sy: *
39: em_hgt_hum: *
40: em_hgt_mus: *
41: em_hgtg_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	100.0	31	6	AX036751	AX036751 Sequence
2	22.6	72.9	152846	2	AC128051	AC128051 Rattus no
3	22	71.0	170897	9	AC020718	AC020718 Homo sapi
4	22	71.0	186665	2	AC101918	AC101918 Mus muscu
5	22	71.0	201646	2	AC116325	AC116325 Mus muscu
6	21.6	69.7	140855	2	AC094340	AC094340 Rattus no
7	21.4	69.0	120953	2	AC124807	AC124807 Mus muscu
8	21.4	69.0	120953	2	AC102902	AC102902 Mus muscu
9	21.4	69.0	194106	2	AC11085	AC11085 Mus muscu
10	21.4	69.0	346897	1	AP002995	AP002995 Mesorhizo
11	21.2	68.4	156159	2	AC023325	AC023325 Homo sapi
12	21.2	68.4	157533	9	AC092868	AC092868 Homo sapi
13	21.2	68.4	189959	9	AC092756	AC092756 Homo sapi
14	21	67.7	106755	2	AL138779	AL138779 Homo sapi
15	21	67.7	129121	2	AC120343	AC120343 Mus muscu
16	21	67.7	134011	9	AL354883	AL354883 Human DNA
17	21	67.7	199160	2	AC110136	AC110136 Rattus no
18	20.8	67.1	191453	2	AL840639	AL840639 Mus muscu
19	20.6	66.5	144014	2	AC098695	AC098695 Oryza sat
20	20.6	66.5	159702	9	AL356433	AL356433 Human DNA
21	20.6	66.5	165423	2	AC008111	AC008111 Homo sapi
22	20.6	66.5	167042	9	AC117527	AC117527 Homo sapi
23	20.6	66.5	170216	2	AC012315	AC012315 Homo sapi
24	20.6	66.5	186932	2	AC113061	AC113061 Mus muscu
25	20.6	66.5	233606	2	AL831778	AL831778 Mus muscu
26	20.4	65.8	18201	3	DVDMAPTP	DVDMAPTP
27	20.4	65.8	26435	3	U00034	U00034 D.vitilis D
28	20.4	65.8	63955	3	AC108151	AC108151 Homo sapi
29	20.4	65.8	78413	2	AC006497	AC006497 Drosophila
30	20.4	65.8	87780	3	AC006714	AC006714 Caenorhab
31	20.4	65.8	92013	2	AC111814	AC111814 Rattus no
32	20.4	65.8	111547	2	AC094583	AC094583 Rattus no
33	20.4	65.8	111967	2	AC010029	AC010029 Drosophila
34	20.4	65.8	124040	2	AC097412	AC097412 Rattus no
35	20.4	65.8	136186	2	AP003896	AP003896 Oryza sat
36	20.4	65.8	141320	2	AC017833	AC017833 Drosophila
37	20.4	65.8	143687	9	AC016572	AC016572 Homo sapi
38	20.4	65.8	147570	2	AC091874	AC091874 Homo sapi
39	20.4	65.8	151514	2	AC113827	AC113827 Rattus no
40	20.4	65.8	153267	2	AC101950	AC101950 Mus muscu
41	20.4	65.8	163191	2	AC107585	AC107585 Rattus no
42	20.4	65.8	169069	2	AC016287	AC016287 Homo sapi
43	20.4	65.8	171612	2	AC068437	AC068437 Homo sapi
44	20.4	65.8	173381	2	AC109087	AC109087 Rattus no
45	20.4	65.8	174805	2	AC107678	AC107678 Mus muscu

ALIGNMENTS

RESULT 1
LOCUS AX036751 31 bp DNA
DEFINITION Sequence 17 from Patent WO0058485.
ACCESSION AX036751
VERSION AX036751.1 GI:11226260
KEYWORDS
ORGANISM
SOURCE
synthetic construct.
synthetic construct.
artificial sequences.
REFERENCE
1 (bases 1 to 31)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 17 05-OCT-2000;

Pred. No. is the number of results predicted by chance to have a

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
 ; GRUBER VERONIQUE (FR)
 FEATURES location/Qualifiers
 source 1. 31
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Guide desoxynucleotide building block G3"
 BASE COUNT 5 a 11 c 2 g 13 t
 ORIGIN
 Query Match 100.0%; Score 31; DB 6; Length 31;
 Best Local Similarity 100.0%; Fred. No. 0.017;
 Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 ATTCACCTTTACATTCCTCCCTTCATG 31
 1 ATTCACCTTTACATTCCTCCCTTCATG 31
 RESULT 2
 AC128051/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-453C4, *** SEQUENCING IN PROGRESS
 ACCESSION AC128051 152846 bp DNA linear HTG 19-JUL-2002
 VERSION AC128051.1 GI:21908645
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 152846)
 Muzny,D.M., Adams,C., Adio-Ogunola,B., Allisonman,F.R., Allen,C.,
 Alebrooke,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbarella,J., Benton,J., Bimaga,K., Blanchenburg,K., Bonnin,D.,
 Bouck,J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dachorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
 Gabibel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Hollaway,C., Hollins,B.,
 Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,U., Liu,W., Lonsleged,H.,
 Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,B.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabac,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokemko,S., Ogulu,M., Okwundu,G.,
 Otagunye,N., Oyedero,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojlobokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
 Sodergren,E., Sonakke,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Umanai,K., Vazquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
 Williams,G., Williamson,A., Wleczek,R., Woodson,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLES
 JOURNAL
 COMMENT
 Direct Submission
 Unpublished
 2 (bases 1 to 152846)
 Morley,K.C.
 Direct Submission
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: KAPL
 Center clone name: CH230-453C4
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 107637 bases at least Q40
 Consensus quality: 112100 bases at least Q30
 Consensus quality: 114998 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 43 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 1146: contig of 1146 bp in length
 1147 1246: gap of unknown length
 1247 2500: contig of 1254 bp in length
 2501 2600: gap of unknown length
 2601 3922: contig of 1322 bp in length
 3923 4022: gap of unknown length
 4023 5151: contig of 1128 bp in length
 5152 5251: gap of unknown length
 5252 6455: contig of 1204 bp in length
 6456 6555: gap of unknown length
 6556 8208: contig of 1653 bp in length
 8209 8309: gap of unknown length
 8309 9596: contig of 1288 bp in length
 9597 9696: gap of unknown length
 9697 11091: contig of 1395 bp in length
 11092 11191: gap of unknown length
 11192 12812: contig of 1621 bp in length
 12813 12912: gap of unknown length
 12913 14030: contig of 1118 bp in length
 14031 14130: gap of unknown length
 14131 15657: contig of 1527 bp in length
 15658 15757: gap of unknown length
 15758 17284: contig of 1527 bp in length
 17285 17385: gap of unknown length
 17385 19260: contig of 1876 bp in length
 19261 19360: gap of unknown length
 19361 21188: contig of 1828 bp in length
 21189 21288: gap of unknown length
 21289 22693: contig of 1405 bp in length
 22694 22794: gap of unknown length
 22794 24286: contig of 1493 bp in length
 24287 24386: gap of unknown length
 24387 26214: contig of 1828 bp in length
 26215 26314: gap of unknown length
 26315 28811: contig of 2497 bp in length
 28812 28912: gap of unknown length
 28912 30769: contig of 1855 bp in length
 30770 30869: gap of unknown length
 30870 34139: contig of 3270 bp in length


```

* 34140 34239: gap of unknown length
* 34240 36862: contig of 2623 bp in length
* 36863 36963: gap of unknown length
* 36963 39221: contig of 2259 bp in length
* 39222 39321: gap of unknown length
* 39321 42654: contig of 3333 bp in length
* 42655 42754: gap of unknown length
* 42755 45445: contig of 2691 bp in length
* 45446 45545: gap of unknown length
* 45546 49130: contig of 3585 bp in length
* 49131 49230: gap of unknown length
* 49231 51555: contig of 2325 bp in length
* 51556 54244: gap of unknown length
* 54245 54344: gap of unknown length
* 54345 58069: contig of 3725 bp in length
* 58070 58170: gap of unknown length
* 58170 62125: contig of 3956 bp in length
* 62126 65502: gap of unknown length
* 65503 70035: gap of unknown length
* 70036 70136: contig of 4434 bp in length
* 70137 73239: gap of unknown length
* 73240 73339: contig of 3103 bp in length
* 73340 77890: gap of unknown length
* 77891 85351: contig of 7361 bp in length
* 85352 85451: gap of unknown length
* 85452 89326: contig of 3875 bp in length
* 89327 95931: contig of 6505 bp in length
* 95932 96031: gap of unknown length
* 96032 101741: contig of 5710 bp in length
* 101742 101841: gap of unknown length
* 101842 108438: contig of 6597 bp in length
* 108439 108538: gap of unknown length
* 108539 115412: contig of 6874 bp in length
* 115413 124473: gap of unknown length
* 124474 124573: gap of unknown length
* 124574 133635: contig of 9061 bp in length
* 133636 133734: gap of unknown length
* 133735 143824: contig of 10090 bp in length
* 143825 143925: gap of unknown length
* 143926 152846: contig of 8922 bp in length.

```

```

FEATURES
  source
    1. 152846
        /organism="Rattus norvegicus"
        /db_xref="taxon:10116"
        /clone="CH230-453C4"

```

```

BASE COUNT  43432 a 30075 c 30565 g 44509 t 4265 others
ORIGIN

```

```

Query Match      72.9%: Score 22.6; DB 2; Length 152846;
Best Local Similarity 86.2%; Pred. No. 47;
Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY  2  TTCCACCTTTTACATTCCTCCGCTTTCATT 30
    |||||
Db 36135 TTCCACCTTTCATTCCTCCCTTTCCTT 36107

```

```

RESULT 3
LOCUS      AC020718      170897 bp      DNA      linear      PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone Rpl1-472L21 from 2, complete sequence.
ACCESSION  AC020718
VERSION     AC020718.7  GI:15383874
KEYWORDS   HTG.
SOURCE      Homo sapiens.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS      Sulston, J.E. and Watson, R.
TITLE        Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE      99063792
PUBMED      9847074
REFERENCE
AUTHORS      Cordes, M., Haakena, W. and Bernard, K.
TITLE        The sequence of Homo sapiens BAC clone Rpl1-472L21
JOURNAL      Unpublished (2001)
REFERENCE
AUTHORS      Watson, R.H.
TITLE        Direct Submission
JOURNAL      Submitted (08-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE
AUTHORS      Watson, R.H.
TITLE        Direct Submission
JOURNAL      Submitted (31-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE
AUTHORS      Watson, R.
TITLE        Direct Submission
JOURNAL      Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            (bases 1 to 170897)
REFERENCE
AUTHORS      Watson, R.
TITLE        Direct Submission
JOURNAL      Submitted (01-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Aug 31, 2001 this sequence version replaced gi:14030137.
            ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            ----- Summary Statistics
            Center project name: H_NH0472L21

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frangen, E., Tateo, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-32011. Actual start of this clone is at base position 1 of RP11-472L21; actual end is at base position 170897 of RP11-472L21.

FEATURES

```
source
1..170897
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-472L21"
/clone_1fb="RPC1-11"
1034..2872
/rpt_family="L1"
1427..1446
/rpt_family="(CAA)n"
2876..3187
/rpt_family="Alu"
3164..3320
/rpt_family="A-rich"
3217..3958
/rpt_family="L1"
3959..4230
/note="match to EST AI445609 (NID:94289409) t308a.x1"
3959..4218
/note="match to EST AW969647 (NID:98159491)"
misc_feature
6425..6565
/rpt_family="Alu"
6546..6565
/rpt_family="(A)n"
6567..7393
/rpt_family="L1"
7390..7424
/rpt_family="GA-rich"
7425..7603
/rpt_family="(GAA)n"
7755..8061
/rpt_family="Alu"
8039..8061
/rpt_family="(A)n"
8844..9064
/rpt_family="L1"
9226..9897
/rpt_family="BRVL"
10497..10604
/rpt_family="GA-rich"
10842..10932
/rpt_family="AT-rich"
11129..11595
/rpt_family="MaLR"
11777..11859
/rpt_family="MIR"
11880..12049
/rpt_family="MIR"
13751..13779
/rpt_family="AT-rich"
13753..13811
/rpt_family="L2"
15421..15635
/rpt_family="MaLR"
15642..15740
/rpt_family="Alu"
15868..15898
/rpt_family="AT-rich"
1601B..16204
/rpt_family="MIR"
16465..16545
/rpt_family="L2"
16632..16699
/rpt_family="MIR"
16819..17027
/rpt_family="L2"
17028..17236
/rpt_family="MER1_type"
```

```
Repeat_region      17237..17319
                    /rpt_family="L2"
Repeat_region      17403..17430
                    /rpt_family="AT_rich"
Repeat_region      18629..18670
                    /rpt_family="AT_rich"
Repeat_region      19324..19449
                    /rpt_family="L2"
Repeat_region      19726..20081
                    /rpt_family="L1"
Repeat_region      19753..19799
                    /rpt_family="(CATA)n"
Repeat_region      21176..21347
                    /rpt_family="MIR"
Repeat_region      21770..21881
                    /rpt_family="MAR"
Repeat_region      21889..21908
                    /rpt_family="(TTTTA)n"
Repeat_region      21892..22185
                    /rpt_family="Alu"
Repeat_region      22186..22286
                    /rpt_family="MAR"
Repeat_region      22396..22558
                    /rpt_family="L2"
Repeat_region      23067..23213
                    /rpt_family="MER1_type"
Repeat_region      23216..23310
                    /rpt_family="MER1_type"
Repeat_region      23327..23618
                    /rpt_family="MAR"
Repeat_region      23619..23860
                    /rpt_family="MER2_type"
Repeat_region      23851..23871
                    /rpt_family="MAR"
Repeat_region      23888..24290
                    /rpt_family="MER1_type"
Repeat_region      24122..24173
                    /rpt_family="AT_rich"
Repeat_region      24405..25001
                    /rpt_family="AT_rich"
Repeat_region      25356..25413
                    /rpt_family="(TTATA)n"
Repeat_region      25436..25512
                    /rpt_family="MIR"
Repeat_region      26616..26689
                    /rpt_family="AT_rich"
```

Query Match	71.0%;	Score 22;	DB 9;	Length 170897;
Best Local Similarity	83.3%;	Pred. No. 83;		
Matches 25; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

Qy 2 TTGCACCTTTTACATTCCGCGCTTTCATTG 31
Db 38430 TTGCACCGTTTACATTCCCACTTTCATTAG 38455

RESULT 4	
AC101918	
LOCUS	
DEFINITION	AC101918 186665 bp DNA linear HTG 21-AUG-2002
ACCSSION	Mus musculus clone RP24-49513, WORKING DRAFT SEQUENCE, 20 unordered
VERSION	pieces.
KEYWORDS	AC101918 AC101918 GI:22381410
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
ORGANISM	house mouse.
	Mus musculus.
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 186665)
TITLE	Birren,B., Nusbaum,C. and Landier,E.
JOURNAL	Mus musculus, clone RP24-49513
REFERENCE	unpublished
	2 (bases 1 to 186665)

AUTHORS

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boughalter,B., Brown,A., Camarata,J., Campolano,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearliano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Harford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamasares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J., Menes,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Vriel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

REFERENCE

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 186665)

JOURNAL

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 186665)

AUTHORS

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boughalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Dearliano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrum,J., Menes,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Vriel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Submitted (21-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Aug 21, 2002 this sequence version replaced gi:17060694.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L17743
Center clone name: 495_1_3

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 180378 bases at least Q40
Consensus quality: 183287 bases at least Q30
Consensus quality: 184280 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 184765; sum-of-ctnigs
Quality coverage: 6.6 in Q20 bases; agarose-fp
Quality coverage: 6.7 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 ctnigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N. The exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1081: contig of 1081 bp in length
1082 1181: gap of 100 bp
1182 2119: contig of 938 bp in length
2120 2219: gap of 100 bp
2220 3423: contig of 1204 bp in length
3424 3523: gap of 100 bp
3524 4709: contig of 1186 bp in length
4710 4809: gap of 100 bp
4810 6162: contig of 1353 bp in length
6163 6262: gap of 100 bp
6263 7591: contig of 1328 bp in length
7592 7691: gap of 100 bp
7692 9907: contig of 2216 bp in length
9908 10007: gap of 100 bp
10008 12103: contig of 2096 bp in length
12104 12203: gap of 100 bp
12204 16361: contig of 4158 bp in length
16362 16461: gap of 100 bp
16462 22485: contig of 6024 bp in length
22486 22585: gap of 100 bp
22586 30730: contig of 8145 bp in length
30731 30830: gap of 100 bp
30831 40176: contig of 9346 bp in length
40177 40276: gap of 100 bp
40277 50333: contig of 10117 bp in length
50394 50493: gap of 100 bp
50494 63206: contig of 12713 bp in length
63207 63306: gap of 100 bp
63307 77181: contig of 13875 bp in length
77182 77281: gap of 100 bp
77282 90693: contig of 13412 bp in length
90694 90793: gap of 100 bp
90794 105614: contig of 14821 bp in length
105615 105714: gap of 100 bp
105715 122131: contig of 16417 bp in length
122132 122231: gap of 100 bp
122232 143060: contig of 20829 bp in length
143061 143160: gap of 100 bp
143161 186665: contig of 43505 bp in length.

FEATURES

source

misc_feature
1..1081
/note="assembly_fragment"
misc_feature
1182..2119
/note="assembly_fragment"
misc_feature
2220..3423
/note="assembly_fragment"
misc_feature
3524..4709
/note="assembly_fragment"
misc_feature
4810..6162
/note="assembly_fragment"
misc_feature
6263..7591
/note="assembly_fragment"
misc_feature
7692..9907
/note="assembly_fragment"
misc_feature
10008..12103
/note="assembly_fragment"
misc_feature
12204..16361
/note="assembly_fragment"
misc_feature
16462..22485
/note="assembly_fragment"
misc_feature
22586..30730
/note="assembly_fragment"

misc_feature 30831..40176 /note="assembly_fragment"
misc_feature 40277..50393 /note="assembly_fragment"
misc_feature 50494..63206 /note="assembly_fragment"
misc_feature 63307..77181 /note="assembly_fragment"
misc_feature 77282..90693 /note="assembly_fragment"
misc_feature 90794..105614 /note="assembly_fragment"
misc_feature 105715..122131 /note="assembly_fragment"
misc_feature 122232..143060 /note="assembly_fragment"
misc_feature 143161..186665 /note="assembly_fragment"
BASE COUNT 53537 a 35894 c 36846 g 58485 t 1903 others
ORIGIN

Query Match 71.0%; Score 22; DB 2; Length 186665;
Best Local Similarity 83.3%; Pred. No. 83;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 TTCGACCTTTACATCCCGCTTCATG 31
Db 96252 TTCACCTTTACTTCTGCTTCCTCACTG 96281

RESULT 5
AC116325/c 201646 bp DNA linear HTG 25-MAY-2002
LOCUS Mus musculus chromosome UNK clone RP23-295K3, WORKING DRAFT
DEFINITION
AC116325
SEQUENCE 9 unordered pieces.
VERSION AC116325.3 GI:21206446
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS 1 (bases 1 to 201646)
McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 201646)
McPherson,J.D. and Waterston,R.H.
AUTHORS Direct Submission
TITLE Submitted (27-MAR-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 201646)
McPherson,J.D. and Waterston,R.H.
REFERENCE Direct Submission
TITLE Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park
JOURNAL Parkway, St. Louis, MO 63108, USA
On May 25, 2002 this sequence version replaced gi:20986680.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
Project Information
Center project name: M_BA0295K03

----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-terminator Big Dye, 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 198905 bases at least Q40

Consensus quality: 199478 bases at least Q30
Consensus quality: 200252 bases at least Q20
Insert size: 203000; agarose-fp
Insert size: 201820; sum-of-contigs
Quality coverage: 14.37 in Q20 bases; agarose-fp
Quality coverage: 12.44 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 6967: contig of 6967 bp in length
* 6968 7067: gap of unknown length
* 7068 14958: contig of 7891 bp in length
* 14959 15058: gap of unknown length
* 15059 24941: contig of 9883 bp in length
* 24942 25042: gap of unknown length
* 25042 36089: contig of 11046 bp in length
* 36090 36190: gap of unknown length
* 36190 50440: contig of 14251 bp in length
* 50441 50541: gap of unknown length
* 50541 85658: contig of 35118 bp in length
* 85659 85758: gap of unknown length
* 85759 145395: contig of 59637 bp in length
* 145396 145495: gap of unknown length
* 145496 201444: contig of 55949 bp in length
* 201445 201545: gap of unknown length
* 201545 201545: contig of 102 bp in length.
Location/Qualifiers
1..201646
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-295K3"
1..6967
/note="assembly_name:Contig24"
7068..14958
/note="assembly_name:Contig25"
15059..24941
/note="assembly_name:Contig26"
25042..36089
/note="assembly_name:Contig27"
36190..50440
/note="assembly_name:Contig28"
50541..85658
/note="assembly_name:Contig29"
85759..145395
/note="assembly_name:Contig30"
145496..201444
/note="assembly_name:Contig31"
201545..201646
/note="assembly_name:Contig39"

FEATURES
source
1..201646
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-295K3"
1..6967
/note="assembly_name:Contig24"
7068..14958
/note="assembly_name:Contig25"
15059..24941
/note="assembly_name:Contig26"
25042..36089
/note="assembly_name:Contig27"
36190..50440
/note="assembly_name:Contig28"
50541..85658
/note="assembly_name:Contig29"
85759..145395
/note="assembly_name:Contig30"
145496..201444
/note="assembly_name:Contig31"
201545..201646
/note="assembly_name:Contig39"

BASE COUNT 61090 a 39295 c 39387 g 61070 t 804 others
ORIGIN

Query Match 71.0%; Score 22; DB 2; Length 201646;
Best Local Similarity 83.3%; Pred. No. 83;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ATTGACCTTTACATCCCGCTTCATG 30
Db 55868 ATTGACGCTTACATTCGCTTCAT 55839

RESULT 6
AC094340 140855 bp DNA linear HTG 10-JUN-2002
LOCUS Rattus norvegicus clone CH230-3J16, *** SEQUENCING IN PROGRESS ***
DEFINITION
78 unordered pieces.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AC094340
AC094340.3 GI:21716583
HTG; HTGS PHASE1.
Norway rat.
Rattus norvegicus
Bakayote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 140855)
Mizny, D.M., Adams, C., Adio-Ochola, B., Ali-Osman, F.R., Allen, C.,
Alshbrook, S.L., Amaralunga, H.C., Are, J.R., Ayale, M., Banks, T.,
Barbala, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Bouch, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earmhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Falle, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Franz, P.,
Gabris, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gottell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsbom, E., Kelly, S., Khan, U., King, L., Korvan, U., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Loulsegged, H.,
Lozad, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Monabadi, K., Morgan, M., Morris, S.,
Mosser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokwuo, S., Ogunu, M., Okunodu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G.,
Scheyer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
Sodergren, E., Sonake, T., Sparke, A., Stanley, H., Stone, H.,
Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tatney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Umami, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

TITLE
JOURNALREFERENCE
AUTHORS

2 (bases 1 to 140855)
Worley, K.C.
Direct Submission
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA.

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 140855)
Worley, K.C.
Direct Submission
Submitted (10-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA.
On Jul 9, 2002 this sequence version replaced gi:17941061.

COMMENT

Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center Project name: GAKM
Center clone name: CH230-3J16
Summary Statistics

Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 58070 bases at least Q40
Consensus quality: 65587 bases at least Q30
Consensus quality: 71764 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 78 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1
1092: contig of 1091 bp in length
1092: gap of unknown length
1192: contig of 1188 bp in length
2379: gap of unknown length
2380: gap of unknown length
2479: gap of unknown length
2480: gap of unknown length
3754: gap of unknown length
3854: gap of unknown length
5013: contig of 1160 bp in length
5113: gap of unknown length
5114: gap of unknown length
5114: contig of 1057 bp in length
6171: gap of unknown length
6271: contig of 1003 bp in length
7271: gap of unknown length
7374: gap of unknown length
8441: gap of unknown length
8441: gap of unknown length
9622: contig of 1082 bp in length
9623: gap of unknown length
9723: gap of unknown length
10887: contig of 1165 bp in length
10887: gap of unknown length
10987: gap of unknown length
12161: contig of 1174 bp in length
12162: gap of unknown length
12262: contig of 1117 bp in length
13378: gap of unknown length
13379: gap of unknown length
13478: gap of unknown length
13479: gap of unknown length
14576: gap of unknown length
14676: gap of unknown length
15862: contig of 1186 bp in length
15863: gap of unknown length
15964: contig of 1002 bp in length
16964: gap of unknown length
17065: gap of unknown length
18127: contig of 1063 bp in length
18128: gap of unknown length
18227: gap of unknown length
19346: contig of 1119 bp in length
19347: gap of unknown length
19446: gap of unknown length
19447: contig of 1028 bp in length
20474: gap of unknown length
20475: gap of unknown length
20574: gap of unknown length
21588: contig of 1015 bp in length
21590: gap of unknown length
21689: gap of unknown length
22820: contig of 1131 bp in length
22821: gap of unknown length
22921: contig of 1015 bp in length
22921: gap of unknown length
23935: gap of unknown length
24035: gap of unknown length
24036: contig of 1051 bp in length
25086: gap of unknown length
25087: gap of unknown length
25186: gap of unknown length
26432: contig of 1246 bp in length
26433: gap of unknown length
26532: gap of unknown length
26533: contig of 1157 bp in length
27689: gap of unknown length
27789: gap of unknown length
27790: contig of 1019 bp in length
28808: gap of unknown length
28809: gap of unknown length
28908: gap of unknown length
30260: contig of 1352 bp in length
30360: gap of unknown length
30361: gap of unknown length
31467: contig of 1107 bp in length
31468: gap of unknown length
31567: gap of unknown length
31568: gap of unknown length
32637: contig of 1070 bp in length
32737: gap of unknown length
32738: contig of 1150 bp in length
33887: gap of unknown length
33888: gap of unknown length

```

* 33988 35644: contig of 1657 bp in length
* 35645 35744: gap of unknown length
* 35745 35806: contig of 1062 bp in length
* 36807 36906: gap of unknown length
* 36907 38037: contig of 1131 bp in length
* 38038 38137: gap of unknown length
* 38138 39567: contig of 1430 bp in length
* 39568 39668: gap of unknown length
* 40801 40901: contig of 1134 bp in length
* 40902 40902: gap of unknown length
* 42348 42348: contig of 1447 bp in length
* 42449 42449: gap of unknown length
* 43984 43984: contig of 1536 bp in length
* 43985 44084: gap of unknown length
* 44085 45144: contig of 1060 bp in length
* 45145 45244: gap of unknown length
* 45245 45245: contig of 1522 bp in length
* 46767 46866: gap of unknown length
* 46867 47868: contig of 1001 bp in length
* 47868 47967: gap of unknown length
* 49241 49241: contig of 1274 bp in length
* 49242 49342: gap of unknown length
* 49343 50451: contig of 1110 bp in length
* 50452 50551: gap of unknown length
* 50552 52685: contig of 2133 bp in length
* 52685 52785: gap of unknown length
* 52785 54915: contig of 2131 bp in length
* 54916 55015: gap of unknown length
* 55016 56201: contig of 1186 bp in length
* 56202 56301: gap of unknown length
* 56302 57981: contig of 1680 bp in length
* 57982 58081: gap of unknown length
* 58082 59684: contig of 1603 bp in length
* 59685 59784: gap of unknown length
* 59785 61419: contig of 1635 bp in length
* 61420 61519: gap of unknown length
* 61520 62778: contig of 1259 bp in length
* 62779 62878: gap of unknown length
* 62879 63942: contig of 1064 bp in length
* 63943 64042: gap of unknown length
* 64043 65840: contig of 1798 bp in length
* 65841 65940: gap of unknown length
* 65941 67504: contig of 1564 bp in length
* 67505 67604: gap of unknown length
* 67605 69282: contig of 1678 bp in length
* 69283 69383: gap of unknown length
* 69383 70606: contig of 1224 bp in length
* 70607 70706: gap of unknown length
* 70707 73687: contig of 2981 bp in length
* 73688 73787: gap of unknown length
* 73788 76196: contig of 2409 bp in length

```

```

Query Match      69.7%  Score 21.6; DB 2; Length 140855;
Best Local Similarity 85.7%  Pred. No. 1.2e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Oy 3 TCCACCTTTTACATCCGCCCTTCAT 30
Db 54797 TCCACCTTTTCCCATCCGCCCTTTCTTT 54824

```

```

RESULT 7
AC124807/c 71489 bp DNA linear HTG 17-JUN-2002
LOCUS AC124807
DEFINITION Mus musculus clone RP23-137113, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC124807
VERSION AC124807.1 GI:21431243
KEYWORDS HTG; HTGS PHASED.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 71489)

```

AUTHORS JOURNAL TITLE REFERENCE AUTHORS

TITLE JOURNAL COMMENT

```

Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-137113
Unpublished
2 (bases 1 to 71489)
Anderson,S., Birtten,B., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dekrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Lacroque,K.,
Lamazeres,R., Landers,T., Lehoccky,J., Levine,R., Lindblad-Toh,K.,
Lu,G., Maclean,C., Macdonald,P., Major,J., Margulis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thoman,N.,
Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Trigglio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26859
Center clone name: 137_I_13
-----
* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 758 857: gap of 100 bp
* 858 1599: contig of 742 bp in length
* 1600 1699: gap of 100 bp
* 1700 2430: contig of 731 bp in length
* 2431 2530: gap of 100 bp
* 2531 3268: contig of 738 bp in length
* 3269 3368: gap of 100 bp
* 3369 4117: contig of 749 bp in length
* 4118 4217: gap of 100 bp
* 4218 4978: contig of 761 bp in length
* 4979 5078: gap of 100 bp
* 5079 5820: contig of 742 bp in length
* 5821 5920: gap of 100 bp
* 5921 6676: contig of 756 bp in length
* 6677 6776: gap of 100 bp
* 6777 7540: contig of 764 bp in length
* 7541 7640: gap of 100 bp
* 7641 8387: contig of 747 bp in length
* 8388 8487: gap of 100 bp
* 8488 9229: contig of 742 bp in length

```

```

* 9230 9329: gap of 100 bp
* 9330 10077: contig of 748 bp in length
* 10078 10177: gap of 100 bp
* 10178 10933: contig of 756 bp in length
* 10934 11033: gap of 100 bp
* 11034 11754: contig of 721 bp in length
* 11755 11854: gap of 100 bp
* 11855 12598: contig of 744 bp in length
* 12599 12698: gap of 100 bp
* 12699 13451: contig of 753 bp in length
* 13452 13551: gap of 100 bp
* 13552 14310: contig of 759 bp in length
* 14311 14410: gap of 100 bp
* 14411 15170: contig of 760 bp in length
* 15171 15270: gap of 100 bp
* 15271 16005: contig of 736 bp in length
* 16007 16106: gap of 100 bp
* 16107 16828: contig of 722 bp in length
* 16829 16928: gap of 100 bp
* 16929 17653: contig of 725 bp in length
* 17654 17753: gap of 100 bp
* 17754 18494: contig of 741 bp in length
* 18495 18594: gap of 100 bp
* 18595 19338: contig of 744 bp in length
* 19339 19438: gap of 100 bp
* 19439 20169: contig of 731 bp in length
* 20170 20269: gap of 100 bp
* 20270 21030: contig of 761 bp in length
* 21031 21130: gap of 100 bp
* 21131 21886: contig of 756 bp in length
* 21887 21986: gap of 100 bp
* 21987 22734: contig of 748 bp in length
* 22735 22834: gap of 100 bp
* 22835 23598: contig of 765 bp in length
* 23600 23699: gap of 100 bp
* 23700 24455: contig of 756 bp in length
* 24456 24555: gap of 100 bp
* 24556 25306: contig of 751 bp in length
* 25307 25406: gap of 100 bp
* 25407 25146: contig of 740 bp in length
* 26147 26246: gap of 100 bp
* 26247 26989: contig of 743 bp in length
* 26990 27089: gap of 100 bp
* 27090 27846: contig of 757 bp in length
* 27847 27946: gap of 100 bp
* 27947 28627: contig of 681 bp in length
* 28628 28727: gap of 100 bp
* 28728 29469: contig of 742 bp in length
* 29470 29569: gap of 100 bp
* 29570 30329: contig of 760 bp in length
* 30330 30429: gap of 100 bp
* 30430 31184: contig of 755 bp in length
* 31185 31284: gap of 100 bp
* 31285 32037: contig of 753 bp in length
* 32038 32137: gap of 100 bp
* 32138 32893: contig of 756 bp in length
* 32894 32993: gap of 100 bp
* 32994 33741: contig of 748 bp in length
* 33742 33841: gap of 100 bp
* 33842 34573: contig of 732 bp in length
* 34574 34673: gap of 100 bp
* 34674 35394: contig of 721 bp in length
* 35395 35494: gap of 100 bp
* 35495 36212: contig of 718 bp in length
* 36213 36312: gap of 100 bp
* 36313 37065: contig of 753 bp in length
* 37066 37165: gap of 100 bp
* 37166 37887: contig of 722 bp in length
* 37888 37987: gap of 100 bp
* 37988 38722: contig of 735 bp in length
* 38723 38822: gap of 100 bp
* 38823 39561: contig of 739 bp in length
* 39562 39661: gap of 100 bp

```

```

* 39662 40416: contig of 755 bp in length
* 40417 40516: gap of 100 bp
* 40517 41269: contig of 753 bp in length
* 41270 41369: gap of 100 bp
* 41370 42130: contig of 761 bp in length
* 42131 42230: gap of 100 bp
* 42231 42996: contig of 766 bp in length
* 42997 43096: gap of 100 bp
* 43097 43845: contig of 749 bp in length
* 43846 43945: gap of 100 bp
* 43946 44650: contig of 705 bp in length
* 44651 44750: gap of 100 bp
* 44751 45497: contig of 747 bp in length
* 45498 45597: gap of 100 bp
* 45598 46366: contig of 769 bp in length
* 46367 46466: gap of 100 bp
* 46467 47206: contig of 740 bp in length
* 47207 47306: gap of 100 bp
* 47307 48051: contig of 745 bp in length
* 48052 48151: gap of 100 bp
* 48152 48885: contig of 734 bp in length
* 48886 48985: gap of 100 bp
* 48986 49744: contig of 759 bp in length
* 49745 49844: gap of 100 bp
* 49845 50610: contig of 766 bp in length
* 50611 50710: gap of 100 bp
* 50711 51463: contig of 753 bp in length
* 51464 51563: gap of 100 bp
* 51564 52299: contig of 736 bp in length
* 52300 52399: gap of 100 bp
* 52400 53124: contig of 725 bp in length
* 53125 53224: gap of 100 bp
* 53225 53955: contig of 731 bp in length
* 53956 54055: gap of 100 bp
* 54056 54768: contig of 713 bp in length
* 54769 54868: gap of 100 bp
* 54869 55618: contig of 750 bp in length
* 55619 55718: gap of 100 bp
* 55719 56451: contig of 733 bp in length
* 56452 56551: gap of 100 bp
* 56552 57278: contig of 727 bp in length
* 57279 57378: gap of 100 bp
* 57379 58115: contig of 737 bp in length
* 58116 58215: gap of 100 bp

```

Query Match 69.0%; Score 21.4; DB 2; Length 71489;
 Best Local Similarity 80.6%; Pred. No. 1.5e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 ATTCCACCTTTTACATTCGCCCTTTCATTG 31
 Db 28853 ATTCCACCTTTTACATTCGCCCTTTCATTG 28823

RESULT 8 AC102902

LOCUS
 DEFINITION Mus musculus clone RP24-313G10, WORKING DRAFT SEQUENCE, 4 unordered
 pieces.

ACCESSION AC102902 GI:22325310
 VERSION AC102902.2
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 120953)
 Birren,B., Nusbaum,C. and Lander,E.
 Mus musculus, clone RP24-313G10

REFERENCE
 JOURNAL Unpublished
 2 (bases 1 to 120953)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Boguskiavkiy,L., Boukhgalter,B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Daxellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S., Hago, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazeres, R., Landers, T., Lepoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McNetters, R., Meltrin, J., Menue, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Vei, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 120953)
AUTHORS
Barron, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bachtien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Daxellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S., Hago, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meltrin, J., Menue, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Vei, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
Submitted (20-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL
On Aug 20, 2002 this sequence version replaced g1:17061012.
COMMENT
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WITR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L20056
Center clone name: 313 G.10
Summary Statistics
Sequencing Vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 120411 bases at least Q40
Consensus quality: 120487 bases at least Q30
Consensus quality: 120554 bases at least Q20
Insert size: 115000; agarose-fp
Insert size: 120653; sum-of-ctngs
Quality coverage: 11.1 in Q20 bases; agarose-fp
Quality coverage: 10.6 in Q20 bases; sum-of-ctngs

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 7151: contig of 7151 bp in length
* 7152 7251: gap of 100 bp
* 7252 15261: contig of 8010 bp in length
* 15262 15361: gap of 100 bp
* 15362 73016: contig of 57655 bp in length
* 73017 73116: gap of 100 bp
* 73117 120953: contig of 47837 bp in length.
Location/Qualifiers
1. 120953
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-313G10"
/clone_lib="RP24-313G10"
1. 7151
/note="assembly_fragment"
7252. 15261
/note="assembly_fragment"
15362. 73016
/note="assembly_fragment"
73117. 120953
/note="assembly_fragment"
clone_end:7
vector_side:right"

BASE COUNT 36211 a 22785 c 22795 g 38857 t 301 others

ORIGIN
Query Match 69.0%; Score 21.4; DB 2; Length 120953;
Best Local Similarity 80.6%; Pred. No. 1.5e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATTCACCTTTACATTCCTCCCTTCATTCG 31
Db 105772 ATTTAACTTTACATTCCTCCCTTCATTCG 105802

RESULT 9
AC121085 194106 bp DNA linear HTG 15-MAY-2002
LOCUS Mus musculus clone RP24-494H11, WORKING DRAFT SEQUENCE, 14 ordered pieces.
AC121085 1 GI:20800155
ACCESSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 194106)
Mus musculus, clone RP24-494H11
Unpublished
2 (bases 1 to 194106)

REFERENCE
AUTHORS Barron, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bachtien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A., Cooke, P., Daxellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S., Hago, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazeres, R., Landers, T., Lepoczky, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meltrin, J., Menue, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,

TITLE
JOURNAL
COMMENT

Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strasburg, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Vel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L21146

Center clone name: 494_H_11

Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 189377 bases at least Q40

Consensus quality: 191587 bases at least Q30

Consensus quality: 192278 bases at least Q20

Insert size: 18800; agarose-fp

Insert size: 192806; sum-of-contigs

Quality coverage: 6.0 in Q20 bases; agarose-fp

Quality coverage: 5.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```
1 1278: contig of 1278 bp in length
1279 1378: gap of 100 bp
1379 2686: contig of 1308 bp in length
2687 2786: gap of 100 bp
2787 4430: contig of 1644 bp in length
4431 4530: gap of 100 bp
4531 8486: contig of 3956 bp in length
8487 8586: gap of 100 bp
8587 12856: contig of 4270 bp in length
12857 12956: gap of 100 bp
12957 22063: contig of 9107 bp in length
22064 22163: gap of 100 bp
22164 30661: contig of 8498 bp in length
30662 30761: gap of 100 bp
30762 42347: contig of 11586 bp in length
42348 42447: gap of 100 bp
42448 55768: contig of 13321 bp in length
55769 55868: gap of 100 bp
55869 77586: contig of 21718 bp in length
77587 77686: gap of 100 bp
77687 103032: contig of 25346 bp in length
103033 103132: gap of 100 bp
103133 127989: contig of 24857 bp in length
127990 128089: gap of 100 bp
128090 155766: contig of 27679 bp in length
155769 155868: gap of 100 bp
155869 194106: contig of 38238 bp in length.
```

FEATURES

source

```
1. 194106
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-494H11"
/clone_lib="RP24 Male Mouse BAC"
```

```
misc_feature 1. 1278
/note="assembly_fragment"
misc_feature 1379. 2686
/note="assembly_fragment"
misc_feature 2787. 4430
/note="assembly_fragment"
misc_feature 4531. 8486
/note="assembly_fragment"
misc_feature 8587. 12856
/note="assembly_fragment"
misc_feature 12957. 22063
/note="assembly_fragment"
misc_feature 22164. 30661
/note="assembly_fragment"
misc_feature 30762. 42347
/note="assembly_fragment"
misc_feature 42448. 55768
/note="assembly_fragment"
misc_feature 55869. 77586
/note="assembly_fragment"
misc_feature 77687. 103032
/note="assembly_fragment"
misc_feature 103133. 127989
/note="assembly_fragment"
misc_feature 128090. 155768
/note="assembly_fragment"
misc_feature 155869. 194106
/note="assembly_fragment"

BASE COUNT 54107 a 42575 c 42309 g 53814 t 1301 others
ORIGIN
```

```
Query Match 69.0%; Score 21.4; DB 2; Length 194106;
Best Local Similarity 80.6%; Pred. No. 1.5e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATTCACCTTTTACATTCGCCCTTCATG 31
Db 171991 ATTCACCTCTTACGTTTCATCTTTCATG 172021
```

```
RESULT 10
AP002995/c 346897 bp DNA linear BCT 15-MAY-2001
LOCUS Mesorhizobium loti DNA, complete genome, section 2/21.
DEFINITION AP002995 BA0000012
ACCESSION AP002995.2 GI:114021442
VERSION
KEYWORDS
SOURCE
ORGANISM Mesorhizobium loti (strain:MAF30309) DNA.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
```

```
REFERENCE
1 (sites)
Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,
Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M.,
Takeuchi, C., Yamada, M. and Tabata, S.
```

```
TITLE Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
JOURNAL DNA Res. 7 (6), 331-338 (2000)
MEDLINE 21082930
2 (bases 1 to 346897)
```

```
REFERENCE
1 (bases 1 to 346897)
Kaneko, T.
```

```
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp.
```

```
COMMENT URL: http://www.kazusa.or.jp/rhizobase/
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934
On May 11, 2001 this sequence version replaced gi:11994963.
FEATURES
1. 346897
Location/Qualifiers
```

```

/organism="Mesorhizobium loti"
/strain="MARP30309"
/db_xref="taxon:381"
74..160
/note="trnS-TGA"
/feature="complement (240..1109)"
/feature="gene="ms18587"
/feature="complement (240..1109)"
/feature="gene="ms18587"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48055.1"
/feature="db_xref="GI:14021443"
/feature="translation="MNDQOTGTGTTALPYTVVRIRAGDVANPOAVHKNGTFYHFRFG
AGNVPLIVSNKHTLVGKMLEPFDALDNGSKRYFGPANTFRLTKOPLFIEHPDPTV
DLAIPLOPIVEANRGQKKEI FNLLIGRETIAPDNI QAILHAITSVLMTVPFSGIDME
KNMLPVVRGILAVNLYLTIDYRGELTFAAFSGSGSVFAF FENMLPDGSGGYTM
LAOPVYFLGLVLSHSGPMTATGKIVPVPTSDQIAOTQMLHLYGCVKASRYELVE
TIRAKISEPTASG"
/feature="complement (1118..1657)"
/feature="gene="m110476"
/feature="complement (1118..1657)"
/feature="gene="m110476"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48056.1"
/feature="db_xref="GI:14021444"
/feature="translation="MPISALIGFEKDGQWRGALALAGEAMAQTAEENFRIEYDREB
KNLVERICRSPDSFISQAAMQAIRRRPGMVL IHNRRHMEKIITPGEPKVPQTV
DGSVHAGDVALGDLREMHKLRAACCKNSHAEVKRALIKRYGEALFSTVERALFC
TGCDDRGSPVLEIHKLP RN"
2193..2558
/feature="gene="m110478"
/feature="2193..2558"
/feature="gene="m110478"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48057.1"
/feature="db_xref="GI:14021445"
/feature="translation="MALALGALMKRLLFYVMTLAVTSGCAADYLNMYDTVLASGD
AMTNSLLOTVDPEPNPNSNNTHTIEGQGRITGVVQYRSPISAGGYSNGCPDINSVA
ADSHGCKRSABSRPGAM"
2651..3412
/feature="gene="m110479"
/feature="2651..3412"
/feature="gene="m110479"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="product="sucinoglycan biosynthesis protein; Exo1"
/feature="protein_id="BAB48058.1"
/feature="db_xref="GI:14021446"
/feature="translation="MPLRLVLVVAASALATQFVMHNGSLPEINLONTIKRPNVI
SGVASVLDGPTIEHMGORVFNGLIDAPESAQVCDDAKGFYPCGRRSABALDGLIAS
KPVQAFVMTDRGRFYGDCTRADGSSVAAVMVEHQALDMPYSSNGAQAQAKAKA
AKVGLMIGSFQAPMDRAKSHSDAQPTAPLPATGNPCNINKNISASGERITYHP
GQRYAVATVISQNKGRMFCSEAAVAAGRRSR"
3429..3782
/feature="gene="m110480"
/feature="3429..3782"
/feature="gene="m110480"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48059.1"
/feature="db_xref="GI:14021447"
/feature="translation="MKHVRRAAVALALSHSSGRNVSSVYSSELRKVNIDASQGV
VDAYDNGSCHIDGTPSLYHYGONSHIDLPQSGGGYDGDIGNSHFLTVSGNSA
ELVDYGGGFFPSST"
/feature="complement (5516..5947)"
/feature="gene="m110481"

```

```

CDS
complement (5516..5947)
/feature="gene="m110481"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48060.1"
/feature="db_xref="GI:14021448"
/feature="translation="MNSDRGTGKDPSTGRFLPGHSGFKKVGSRSRVROAIDTANRHA
ATPMMLVGNKMLAGDIRAEILLRYGADDFLPKRRVVELPDVENPISMLATRAVLSAA
WSGEISPDPAKQLDIAIAKGKIEBAELLRRVQLEAVKR"
complement (6065..6883)
/feature="gene="m110482"
/feature="complement (6065..6883)"
/feature="gene="m110482"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48061.1"
/feature="db_xref="GI:14021449"
/feature="translation="MIDPKRIDGASSRGSFLPOMNNGSVISAFMPEVSTMSPDIDA
ASATLNGVLAAGNAVIGAERLKRDEIRDVSLERLARYVAAPRYLADRAIDERN
VIGSRNSLAQPIQIENPTPTSLADSLERQVRNMDGPDQGEAFDMQSGCITVL
NAVAPSIDLTIFSKQQLAGLIVEERAIKRMARTIDNPSNDNLATANPLKNMSDDL
VNDLAKARLDGKLREQAVSDAEKLLAVIARASQAGNLTPDEAFKMLGQKVA"
complement (7133..7753)
/feature="gene="m110483"
/feature="complement (7133..7753)"
/feature="gene="m110483"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48062.1"
/feature="db_xref="GI:14021450"
/feature="translation="MANYVILRLTVMKRASITTDANPAKLGQIADPAFLMSRT
REWASVNPSEFOYRDECDVNGSGEIGNIRITFSFKTYQSTIFMDAATTDGHPV
WGLTGCCADNMVVSSTHFSFGIINGVSASVQSDMWQDESSSTVSSBMLDRHLPW
RDPPTGVAROSKELPPAIRLALVGLYRSKOTRRMTAPNQAFENIH"
complement (7827..10169)
/feature="gene="m110485"
/feature="complement (7827..10169)"
/feature="gene="m110485"
/feature="hypothetical protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48063.1"
/feature="db_xref="GI:14021451"
/feature="translation="MRDSSVRLPELLPYRQFVWRRLDQKRGSPKATMPSYANGCR
ADTNPATWSDYPTAVRALASQYAGIGVFVENDPFPVIDLGCDATYGLSABAA
AITRMPNPAWELSGEGEIGHIFGRCDKVALGFRKMAKMGCFYRTKRFVAFPGGI
OGNIDIDNTNOLAIIIVPARGVGEODASQAVRBDPMIIPDDEBELRFLASTGCTK
AKIOASPTNLOLMEHAAALAOYFSPATKGFBDHSSADQALANVLA FMTGRDQVRIIR
LREKSLGRERKTRRAAYRNSITVNAVAVSCQTVYGCNQGRSRQVQENVRIGSIG
ADITHVWLTLEMDLQDLWFIIRHSKSGALADSRTLNVLPPATDQREYSASRYEVAAS
IDPKTDGFKRRLPVFNQWQSPSRKSVVITWKPNGAVICAPBESQAGFNTWRGFO
PYVLPQSGOSLVVWSEHLYLPIEAEVRFEQWLAHIVQPEELPOSAYLFTIET
GIGRNLSSVLVAVRVAAGIADIDEMLDGTENGRLSKLLAVVEEAAGQGYGRW
SHSEKIKSMINPEMRINKEHGLEFVEMNCBMLPFSNMMDLPPDASRRRVVAVNP
THRQSPDYTERITSQIANPDTISAKRRLBLDISFRKCALPDNRBAKRLGAAAS
DIDNLIAARDEMPGLARSDVINYLHVSIAKLPEKTKINKLIEQAGMKLVERISIG
PKRYRVRVVDYTAQOITNPQMAVAVAANARKGXFRQ"
complement (10162..10392)
/feature="gene="m110486"
/feature="complement (10162..10392)"
/feature="gene="m110486"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48064.1"
/feature="db_xref="GI:14021452"
/feature="translation="MECVAMAGNLVREMNQSAIYLSIPOVCATLGISRPTLDRIYKD
NBSFPRKKIKGPLVGLFMSDIETIYGLSDA"
complement (10595..11803)
/feature="gene="m110487"

```

```

gene
complement (5516..5947)
/feature="gene="m110481"

```

```

CDS
complement (7827..10169)
/feature="gene="m110485"
/feature="hypothetical protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48063.1"
/feature="db_xref="GI:14021451"
/feature="translation="MRDSSVRLPELLPYRQFVWRRLDQKRGSPKATMPSYANGCR
ADTNPATWSDYPTAVRALASQYAGIGVFVENDPFPVIDLGCDATYGLSABAA
AITRMPNPAWELSGEGEIGHIFGRCDKVALGFRKMAKMGCFYRTKRFVAFPGGI
OGNIDIDNTNOLAIIIVPARGVGEODASQAVRBDPMIIPDDEBELRFLASTGCTK
AKIOASPTNLOLMEHAAALAOYFSPATKGFBDHSSADQALANVLA FMTGRDQVRIIR
LREKSLGRERKTRRAAYRNSITVNAVAVSCQTVYGCNQGRSRQVQENVRIGSIG
ADITHVWLTLEMDLQDLWFIIRHSKSGALADSRTLNVLPPATDQREYSASRYEVAAS
IDPKTDGFKRRLPVFNQWQSPSRKSVVITWKPNGAVICAPBESQAGFNTWRGFO
PYVLPQSGOSLVVWSEHLYLPIEAEVRFEQWLAHIVQPEELPOSAYLFTIET
GIGRNLSSVLVAVRVAAGIADIDEMLDGTENGRLSKLLAVVEEAAGQGYGRW
SHSEKIKSMINPEMRINKEHGLEFVEMNCBMLPFSNMMDLPPDASRRRVVAVNP
THRQSPDYTERITSQIANPDTISAKRRLBLDISFRKCALPDNRBAKRLGAAAS
DIDNLIAARDEMPGLARSDVINYLHVSIAKLPEKTKINKLIEQAGMKLVERISIG
PKRYRVRVVDYTAQOITNPQMAVAVAANARKGXFRQ"
complement (10162..10392)
/feature="gene="m110486"
/feature="complement (10162..10392)"
/feature="gene="m110486"
/feature="unknown protein"
/codon_start=1
/feature="table=1"
/feature="protein_id="BAB48064.1"
/feature="db_xref="GI:14021452"
/feature="translation="MECVAMAGNLVREMNQSAIYLSIPOVCATLGISRPTLDRIYKD
NBSFPRKKIKGPLVGLFMSDIETIYGLSDA"
complement (10595..11803)
/feature="gene="m110487"

```

```

gene
complement (10595..11803)
/feature="gene="m110487"

```

CDs complement (10595..11803)
/gene="ml10487"
/codon_start=1
/transl_table=11
/product="probable prophage integrase"
/protein_id="BAB48065.1"
/db_xref="GI:14021453"
/translation="MLTIREIQLVHKGPQSHYKALYFPIAPSGRPFAIRYMLK
GRRRSLTDSIPATVRELKQELSETEBKVRKGIPIGLAEBAKAASPEPVRLQV
TPKSVLDYIAANGGKNAKHVOGMENTATVYVPIGLAEPOSITTYLVKLIQPL
MATKRTASVRSRVERITIGAAKAKGSEPEIWMANHNANRANNDIEFLNKEKN
HFSAMPPEIIPAQOSTAEREDYSTALQPLILFARTNNTLGAENSEIDTEAAVWT
PASRMKSGKEHVPISSEALKVLSLRLIGNPYLPGALRNLSLNMALIMVWRGK
LGHVPHGFRSAFRDMAAEETVYPNHI VEMALAHITIKNKVAAVRRGDLMKRRQJME
DMSAYLASAR"
12686..13279
/gene="mlr0488"
12686..13279
/gene="mlr0488"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB48066.1"
69.0%; Score 21.4; DB 1; Length 346897;
Best Local Similarity 80.6%; Pred. No. 1.5e+02;
Matches 45; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ATTCGACCTTTACATTCGCCCTTTCATTG 31
Db 37212 ATCCGACCTTTGCTTCACGCTTTCATTG 37182
RESULT 11
AC023325 156159 bp DNA linear HTG 14-FEB-2001
LOCUS Homo sapiens chromosome 15 clone RP11-356M20 map 15, WORKING DRAFT
DEFINITION AC023325
SEQUENCE AC023325 15 unordered pieces.
AC023325 3 GI:12658138
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 156159)
2 (bases 1 to 156159)
Britten, B., Linton, L., Nuebaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bede, P., Boguslavsky, L.,
Boukhalter, B., Brown, A., Burkett, G., Campolano, A., Castle, A.,
Choepe, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearlano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
Fenster, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
Galaun, J., Gardyna, S., Ginde, S., Goeyete, M., Graham, L.,
Grand, P., Grant, G., Hago, B., Heaford, A., Horton, L.,
Howland, J., C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatae, A.,
Klein, J., Landers, T., Largocque, K., Lehocsky, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,
McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Mihova, T., Miranda, C., Mianga, V., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M.,
Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Testave, S., Theodore, J., Tirrell, A.,
Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.
Direct Submission
Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 3, 2001 this sequence version replaced gi:7139826.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WMR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 356 M 20
Center clone name: 356 M 20
----- Summary Statistics
Sequencing vector: M13; M77815; 71% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 149772 bases at least Q40
Consensus quality: 152333 bases at least Q30
Consensus quality: 153705 bases at least Q20
Insert size: 15700; agarose-fp
Insert size: 154759; sum-of-contigs
Quality coverage: 6.4 in Q20 bases; agarose-fp
Quality coverage: 6.5 in Q20 b.
NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 9882: contig of 9882 bp in length
9883 9982: gap of 100 bp
9983 11301: contig of 1319 bp in length
11302 11401: gap of 100 bp
11402 12615: contig of 1214 bp in length
12616 12715: gap of 100 bp
12716 16302: contig of 3587 bp in length
16303 16402: gap of 100 bp
16403 18619: contig of 2217 bp in length
18620 18719: gap of 100 bp
18720 21764: contig of 3045 bp in length
21765 21864: gap of 100 bp
21865 66492: contig of 44628 bp in length
66493 66592: gap of 100 bp
66593 69609: contig of 3017 bp in length
69610 69709: gap of 100 bp
69710 75994: contig of 6285 bp in length
75995 76094: gap of 100 bp
76095 82300: contig of 6206 bp in length
82301 82400: gap of 100 bp
82401 91215: contig of 8815 bp in length
91216 91315: gap of 100 bp
91316 101994: contig of 10679 bp in length
101995 102094: gap of 100 bp
102095 128651: contig of 26557 bp in length
128652 128751: gap of 100 bp
128752 153403: contig of 24652 bp in length
153404 153503: gap of 100 bp
153504 156159: contig of 2656 bp in length.
Location/Qualifiers
1..156159
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-356M20"
/clone_lib="RP11-11 Human Male BAC"
1..9882
/note="assembly_fragment
clone_end:SP6

FEATURES

Source

misc_feature
1..9882
clone_end:SP6

```

misc_feature      vector:side:left
                  9983..11301
/note="assembly_fragment"
misc_feature      11402..12615
/note="assembly_fragment"
misc_feature      12716..16302
/note="assembly_fragment"
misc_feature      16403..18619
/note="assembly_fragment"
misc_feature      18720..21764
/note="assembly_fragment"
misc_feature      21865..66492
/note="assembly_fragment"
misc_feature      66593..69609
/note="assembly_fragment"
misc_feature      69710..75994
/note="assembly_fragment"
misc_feature      76095..82300
/note="assembly_fragment"
misc_feature      82401..91215
/note="assembly_fragment"
misc_feature      91316..101994
/note="assembly_fragment"
misc_feature      102095..128651
/note="assembly_fragment"
misc_feature      128752..153403
/note="assembly_fragment"
misc_feature      153504..156159
/note="assembly_fragment"
                  /note="assembly_fragment"
                  clone_end:T7
                  vector:side:right"

```

BASE COUNT 44142 a 33879 c 33433 g 43302 t 1403 others
ORIGIN

Query Match 68.4%; Score 21.2; DB 2; Length 156159;
Best Local Similarity 88.5%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATTCACCTTTACATTCGCCCTTT 26
Db 72877 ATTCACCTTTACATTCGCCATT 72902

RESULT 12 AC092868 157533 bp DNA linear PRI 22-NOV-2001
AC092868 Homo sapiens chromosome 15 clone RP11-356M20 map 15q21.3, complete
LOCUS sequence.
AC092868.2 GI:17047069
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 157533)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dore, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.

TITLE Sequencing of human chromosome 15 DIS146-D15S117 region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 157533)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dore, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.

TITLE Direct Submission
JOURNAL Submitted (01-AUG-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA

REFERENCE 3 (bases 1 to 157533)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dore, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.

TITLE Direct Submission
JOURNAL Submitted (22-NOV-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT On Nov 22, 2001 this sequence version replaced gi:15055287.

----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMSC
Web site: http://chroma.mbr.washington.edu/msg_www
Contact: leerowen@systembiology.org
----- Summary Statistics
Sequencing vector: pUC18; 108752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399

Note: Data from overlapping BACs AC092756 [drafting center:
UMSC], and AC091748 [drafting center: UMSC] were added for
finishing

FEATURES
source Location/Qualifiers

1..157533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q21.3"
/clone="RP11-356M20"
/note="Data from overlapping BACs RP11-429D19 and
RP11-99L18 were added and the consensus sequence was
determined from RP11-356M20 to the extent possible"
1..24730
/note="overlap with RP11-429D19 AC092756"
28322..157533
/note="overlap with RP11-99L18 AC091748"
147556..147559
/note="low quality data"

BASE COUNT 45312 a 34917 c 33839 g 43465 t
ORIGIN

Query Match 68.4%; Score 21.2; DB 9; Length 157533;
Best Local Similarity 88.5%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATTCACCTTTACATTCGCCCTTT 26
Db 13088 ATTCACCTTTACATTCGCCATT 13113

RESULT 13 AC092756 189999 bp DNA linear PRI 19-JAN-2002
AC092756 Homo sapiens chromosome 15 clone RP11-429D19 map 15q21.3, complete
LOCUS sequence.
AC092756
AC092756.2 GI:18249990
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 189999)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dore, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.

TITLE Sequencing of human chromosome 15 DIS146-D15S117 region
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 189999)
AUTHORS Rowen, L., Madan, A., Qin, S., Baradarani, L., Birditt, B., Bloom, S.,
Burke, J., Dore, M., Fleetwood, P., Kaur, A., Madan, A., Nesbitt, R.,
Pate, D. and Hood, L.

TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Multimegabase Sequencing Center, Institute

for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
3 (bases 1 to 189999)
Rowen, L., Madan, A., Qin, S., Bardarani, L., Birditt, B., Bloom, S., Burke, J., Dore, M., Fleetwood, P., Kaur, A., Madan, A., Neebitt, R., Pate, D., and Hood, L.
Direct Submission
Submitted (19-JAN-2002) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
On Jan 19, 2002 this sequence version replaced gi:15021994.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UMWSC
Web site: http://chroma.mbc.washington.edu/mag_www
Contact: leerowen@systembiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-terminator Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Data from overlapping BACs AC092757 [drafting center: UMWSC], AC023890 [drafting center: WUGSC], and AC092868 [drafting center: WIBR] was added for finishing
----- Location/Qualifiers
source 1.189999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q21.3"
/clone="RP11-429D19"
/note="Data from overlapping BACs RP11-59H7, RP11-756K9, and RP11-356M20 was added and the consensus sequence was determined from RP11-429D19 to the extent possible"
misc_feature 1.147389
/note="Overlap with: RP11-756K9 AC023890"
misc_feature 1.56595
/note="Overlap with RP11-59H7 AC092757"
137567
/note="low quality data"
164999
/note="low quality data"
misc_feature 165270.189999
/note="Overlap with RP11-356M20"
BASE COUNT 54028 a 41117 c 41361 g 53493 t
ORIGIN
Query Match 68.4%; Score 21.2; DB 9; Length 189999;
Best Local Similarity 88.5%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATTCACCTTTACATTCGCCCTTT 26
Db 178357 ATTCACCTTTACATTCGCCCTT 178382
RESULT 14
AL138779/c 106755 bp DNA linear HTG 10-JUN-2001
LOCUS Homo sapiens chromosome 1 clone RP6-65F20 map p32.2-34.1, ***
DEFINITION SEQUENCING IN PROGRESS ***, 20 unordered pieces.
ACCESSION AL138779
VERSION AL138779.7 GI:9930811
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106755)
McLay, K.
AUTHORS

TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9926429.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: DA65F20
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 14% of reads
Chemistry: Dye-terminator ET-amersham; 85% of reads
Consensus quality: 99905 bases at least Q40
Consensus quality: 102097 bases at least Q20
Insert size: 104855; sum-of-contigs
Insert size: 117633; 9.0% error; agarose-fp
Quality coverage: 3.06x in Q20 bases; sum-of-contigs Quality coverage: 3.01x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 20 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1
2183 2282: contig of 2182 bp in length
2283 11274: contig of 8992 bp in length
11275 11374: gap of 100 bp
11375 17141: contig of 5767 bp in length
17142 17241: gap of 100 bp
17242 21606: contig of 4365 bp in length
21607 21706: gap of 100 bp
21707 26686: contig of 4980 bp in length
26687 26786: gap of 100 bp
26787 31996: contig of 5210 bp in length
31997 32096: gap of 100 bp
32097 35605: contig of 3509 bp in length
35606 35705: gap of 100 bp
35706 49561: contig of 13856 bp in length
49562 49661: gap of 100 bp
49662 52981: contig of 3320 bp in length
52982 53081: gap of 100 bp
53082 62347: contig of 9266 bp in length
62348 62447: gap of 100 bp
62448 66552: contig of 4105 bp in length
66553 66652: gap of 100 bp
66653 69451: contig of 2799 bp in length
69452 69551: gap of 100 bp
69552 73006: contig of 3455 bp in length
73007 73106: gap of 100 bp
73107 77114: contig of 4008 bp in length
77115 77214: gap of 100 bp
77215 80477: contig of 3263 bp in length
80478 80577: gap of 100 bp
80578 85312: contig of 4735 bp in length
85313 85412: gap of 100 bp
85413 89155: contig of 3743 bp in length
89156 89255: gap of 100 bp
89256 92958: contig of 3703 bp in length
92959 93058: gap of 100 bp
93059 98962: contig of 5904 bp in length
98963 99062: gap of 100 bp
99063 106755: contig of 7693 bp in length.
Location/Qualifiers

```
source
1. .106755
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p32.2-34.1"
/clone="RP6-65F20"
/clone_id="RP6-65F20"
1. .2182
/note="assembly fragment:00605
fragment chain:1"
clone_end:SP6
vector_side:left"
2283. .11274
/note="assembly fragment:00337
fragment chain:1"
11375. .17141
/note="assembly fragment:00999
fragment chain:2"
17242. .21606
/note="assembly fragment:00285
fragment chain:2"
21707. .26686
/note="assembly fragment:00197
fragment chain:2"
26787. .31996
/note="assembly fragment:00295
fragment chain:2"
32097. .35605
/note="assembly fragment:00488
fragment chain:3"
35706. .49561
/note="assembly fragment:00322
fragment chain:3"
49662. .52981
/note="assembly fragment:00261
fragment chain:3"
53082. .62347
/note="assembly fragment:00320
fragment chain:4"
62448. .66552
/note="assembly fragment:00882
fragment chain:4"
66653. .69451
/note="assembly fragment:00943
fragment chain:5"
69552. .73006
/note="assembly fragment:00204
fragment chain:5"
73107. .77114
/note="assembly fragment:00302"
77215. .80477
/note="assembly fragment:00594"
80578. .85312
/note="assembly fragment:00828"
85413. .89155
/note="assembly fragment:00646
fragment chain:6"
89256. .92958
/note="assembly fragment:00058
fragment chain:6"
93059. .98962
/note="assembly fragment:00457
fragment chain:6"
99063. .106755
/note="assembly fragment:00019
fragment chain:6"
clone_end:T7
vector_side:right"

BASE COUNT 30905 a 21750 c 21309 g 30884 t 1907 others
Origin
```

```
Query Match 67.7%; Score 21; DB 2; Length 106755;
Best Local Similarity 82.8%; Pred. No. 2.2e+02;
```

```
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATTGCACCTTTACATCCGCCCTTCAT 29
Db 52888 AATTGACCTTGATCATTCGCTTCAT 52860

RESULT 15
AC120343/c
LOCUS
DEFINITION
MUS musculus clone RP23-149H21, WORKING DRAFT SEQUENCE, 10 ordered
pieces.
ACCESSION
AC120343
VERSION
AC120343.3 GI:22381827
KEYWORDS
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
house mouse.
ORGANISM
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 129121)
Birren, B., Nuebaum, C. and Lander, E.
MUS musculus, clone RP23-149H21
2 (bases 1 to 129121)
Unpublished
2 (bases 1 to 129121)
Anderson, S., Barna, N., Nuebaum, C., Lander, E., Ali, A., Allen, N.,
Boukhalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Deatrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galsgan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karacas, A., Kells, C., Lacroque, K., Lamazares, R.,
Lander, T., Lechoczky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menues, L.,
Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 129121)
Birren, B., Nuebaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Deatrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galsgan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karacas, A., Kells, C., Lander, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Menues, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuppback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:20503130.
All repeats were identified using RepeatMasker:
```

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L19845
 Center clone name: 148 H.21
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 126860 bases at least Q40
 Consensus quality: 127509 bases at least Q30
 Consensus quality: 127930 bases at least Q20
 Insert size: 18200; agarose-fp
 Insert size: 18221; sum-of-contigs
 Quality coverage: 7.2 in Q20 bases; agarose-fp
 Quality coverage: 10.2 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 10 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 *
 1 1147: contig of 1147 bp in length
 * 1148 1247: gap of 100 bp
 * 1248 2769: contig of 1522 bp in length
 * 2770 2869: gap of 100 bp
 * 2870 5607: contig of 2738 bp in length
 * 5608 5707: gap of 100 bp
 * 5708 10061: contig of 4354 bp in length
 * 10062 10161: gap of 100 bp
 * 10162 17248: contig of 7087 bp in length
 * 17249 17348: gap of 100 bp
 * 17349 27373: contig of 10025 bp in length
 * 27374 27473: gap of 100 bp
 * 27474 41686: contig of 14213 bp in length
 * 41687 41786: gap of 100 bp
 * 41787 64402: contig of 22616 bp in length
 * 64403 64502: gap of 100 bp
 * 64503 104541: contig of 40039 bp in length
 * 104542 104641: gap of 100 bp
 * 104642 129121: contig of 24480 bp in length.
 Location/Qualifiers
 1. 129121
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-149H21"
 /clone_lib="RPCI-23 Female Mouse BAC"
 1. 1147
 /note="assembly_fragment"
 1248. 2769
 /note="assembly_fragment"
 2870. 5607
 /note="assembly_fragment"
 5708. 10061
 /note="assembly_fragment"
 10162. 17248
 /note="assembly_fragment"
 17349. 27373
 /note="assembly_fragment"
 27474. 41686
 /note="assembly_fragment"
 41787. 64402
 /note="assembly_fragment"
 64503. 104541
 misc_feature

misc_feature /note="assembly_fragment"
 104642. 129121
 /note="assembly_fragment"
 clone_end:T7
 vector_side:right"
 BASE COUNT 34269 a 28590 g 36358 t 904 others
 ORIGIN
 Query Match 67.7%; Score 21; DB 2; Length 129121;
 Best Local Similarity 82.8%; Pred. No. 2.2e+02;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Db 3 TCCACCTTTACATCCCGCTTCATTG 31.
 46410 TCCACCTTTTCCTTCACCTTCATTG 46382
 Search completed: May 11, 2003, 06:16:37
 Job time : 570.336 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 530.481 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-16

Perfect score: 33

Sequence: 1 gctgtgtctactctctctctctctcaccagca 33

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthm:*
3: em_estln:*
4: em_estmv:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_tod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23.6	71.5	996	17	AL068002 Drosophila
C 2	23.6	71.5	1239	12	BG258346 602379796
C 3	23.2	70.3	619	17	BH085145
C 4	23	69.7	479	12	BF397893
C 5	23	69.7	1522	11	AK005882 Mus muscu
C 6	22.6	68.5	634	14	BQ247773 TAE250426

C	7	22.6	68.5	697	14	BQ249531
C 8	22.6	68.5	1527	12	BE876684	BQ249531 TAE250426
C 9	22.4	67.9	693	17	BE876684	BE876684 601488248
C 10	22.2	67.3	455	17	A0872730	BE876684 BOME9197
C 11	22.2	67.3	473	17	A0874006	A0872730 V15F1 mTh
C 12	22.2	67.3	473	17	A0874006	A0874006 V9G9 mTh
C 13	22.2	67.3	577	17	A0872889	A0872889 V54C5 mTh
C 14	22.2	67.3	708	12	BG850976	A0872952 V56B3 mTh
C 15	22.2	67.3	810	17	A0876136	BG850976 102402980
C 16	22.2	67.3	887	17	A2685547	A0876136 V133F7 mT
C 17	22.2	66.7	545	13	BM270934	A2685547 ENTKE48TR
C 18	21.8	66.1	667	17	A2494068	BM270934 sak03C12
C 19	21.8	66.1	854	12	BF688017	A2494068 1M0339N06
C 20	21.8	66.1	1176	13	BI416154	BF688017 602067220
C 21	21.6	65.5	144	10	BE152640	BI416154 602989006
C 22	21.6	65.5	375	13	BT007043	BE152640 CM1-HT032
C 23	21.6	65.5	429	11	BC008693	BT007043 RCS-RT005
C 24	21.6	65.5	443	10	BB847572	BC008693 Homo BAB1
C 25	21.6	65.5	460	10	AW984336	BB847572 BB847572
C 26	21.6	65.5	464	12	BF706719	AW984336 PM3-HN001
C 27	21.6	65.5	473	10	AW984414	BF706719 281344 MA
C 28	21.6	65.5	499	12	BF439546	AW984414 PM3-HN001
C 29	21.6	65.5	505	9	AI963143	BF439546 nab67G02
C 30	21.6	65.5	509	13	BM511124	AI963143 wt27H04.X
C 31	21.6	65.5	534	9	AI747788	BM511124 Ij44C10.Y
C 32	21.6	65.5	536	12	BF042283	AI747788 u102d07.Y
C 33	21.6	65.5	538	12	BF042283	BG896015 359762 MA
C 34	21.6	65.5	548	12	BE722524	BF042283 BP250003A
C 35	21.6	65.5	563	12	BE835234	BE722524 190803 MA
C 36	21.6	65.5	576	12	BG716076	BE835234 RCS-FN002
C 37	21.6	65.5	578	17	AQ309992	BG716076 602677503
C 38	21.6	65.5	585	14	BQ018935	AQ309992 CITR1-E1-
C 39	21.6	65.5	606	14	BM761228	BQ018935 UI-H-DH1-
C 40	21.6	65.5	644	13	BM510872	BM761228 K-EST0041
C 41	21.6	65.5	653	12	BE885377	BM510872 Ij44C10.X
C 42	21.6	65.5	679	12	BG504516	BE885377 601507456
C 43	21.6	65.5	686	14	BQ019003	BG504516 602552488
C 44	21.6	65.5	688	13	BM551574	BQ019003 UI-H-DH1-
C 45	21.6	65.5	691	14	BM677596	BM551574 AGENCOURT

ALIGNMENTS

RESULT 1
CNS00LES/c 996 bp DNA linear GSS 03-JUN-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC:
BACR30A08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Oono and
Aaron Mammmer in Pletier de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial

ECORI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Source

1. 996
/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="BACR30A08"

/clone_id="RP1-98"

/note="end : 17"

BASE COUNT 387 a 161 c 149 g 222 t 77 others

ORIGIN

Query Match 71.5%; Score 23.6; DB 17; Length 996;
Best Local Similarity 81.2%; Pred. No. 3.9e+03;
Matches 26; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

1 GCTGTTACTTTCTCTCTCTATTCAGCC 32

Db 836 GCTCTTTATTTCTCTCTCTCTGCGAGCC 805

RESULT 2
BG258346/c 1239 bp mRNA linear EST 13-FEB-2001

LOCUS 622379796F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4510703 5,

DEFINITION mRNA sequence.

ACCESSION BG258346

VERSION BG258346.1 GI:12768162

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 1239)

NIH-MGC <http://mgs.nci.nih.gov/>.

Unpublished (1999)

Contact: Robert Straubeberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: LLM10392 row: c column: 24

High quality sequence stop: 423.

Location/Qualifiers

1. 1239

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4510703"

/clone_id="NIH_MGC_92"

/tissue_type="embryonal carcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NciI;

Site 2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 2.5 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH_MGC Library."

BASE COUNT 364 a 366 c 346 g 163 t

ORIGIN

Query Match 71.5%; Score 23.6; DB 12; Length 1239;

Best Local Similarity 86.7%; Pred. No. 3.6e+03;

Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

3 TGTGTTACTTTCTCTCTATTCAGCC 32

Db 835 TGTGTTACTTTATTTCTCTGCGAGCC 806

RESULT 3
BH085145/c 619 bp DNA linear GSS 18-JUL-2001

LOCUS RPI-24-400H14.TV RPI-24 Mus musculus genomic clone RPI-24-400H14

DEFINITION DNA sequence.

ACCESSION BH085145

VERSION BH085145.1 GI:14904742

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 619)

Zhao, S., Mierman, W., Malek, J., Shatsman, S., Akintet, B., Levins, M.,

Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gabregeorgis, E.,

Russell, D., de Jong, P., and Fraser, C.M.

Mouse BAC End Sequences from Library RPI-24

Unpublished (1999)

Other_GSSs: RPI-24-400H14.TU

Contact: Shaoying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPI-24. For BAC

library availability, please contact Pieter de Jong

(pdj@tigr.org).

Clones may be purchased from BACPAC

Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end

page: http://www.chori.org/cdb/bac_end/mouse/bac_end_intro.html

Plate: 400 row: H column: 14

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. 619

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPI-24-400H14"

/clone_id="RPI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARACT; Site 1: BamHI, Site 2: BamHI;

RPI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARACT cloning vector at the

BamHI sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 203 a 100 c 124 g 192 t

ORIGIN

Query Match 70.3%; Score 23.2; DB 17; Length 619;

Best Local Similarity 89.3%; Pred. No. 5.6e+03;

Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2 CTGTTGTTACTTTCTCTCTATTCACA 29

Db 515 CTGTTGTTACTTTCTCTCTATTCACA 488

RESULT 4

BF397893 479 bp mRNA linear EST 27-NOV-2000

LOCUS UI-R-B52-bec-h-04-0-UI.61 UI-R-B52 Ratius norvegicus cDNA clone

DEFINITION UI-R-B52-bec-h-04-0-UI 3', mRNA sequence.

ACCESSION BF397893

VERSION BF397893.1 GI:11382878

KEYWORDS EST.

Norway rat.

ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 479)
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.wieg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence track present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized embryo at 13 dpc library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source location/Qualifiers
 1..479
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BS2-bec-h-04-0-UI"
 /clone_11b="UI-R-BS2"
 /dev_stage="embryonic 13 dpc"
 /lab_host="DHI0B (Life Technologies)"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BS2 library is a subtracted library derived from 13 dpc whole embryo tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratsec.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_L1B=UI-R-BS2
 TAG_TISSUE=embryo at 13 dpc
 TAG_SEQ=ATTC"
 BASE COUNT 145 a 92 c 108 g 134 t
 ORIGIN

Query Match 69.7%; Score 23; DB 12; Length 479;
 Best Local Similarity 83.9%; Pred. No. 6.9e+03;
 Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TGTGTACTTCTCTCTATTCAGCCCA 33
 Db 383 TGTGTCTTTCTTCTCTATTCAGACA 143

RESULT 5
 AK005882/c
 LOCUS
 DEFINITION AK005882 1522 bp mRNA linear HTC 19-JAN-2002
 Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700011N24; Retroviral aspartyl protease containing protein, full insert sequence.
 ACCESSION AK005882
 VERSION AK005882.1 GI:12838694
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain: C57BL/6J) adult male testis cDNA to mRNA, clone:1700011N24.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, Y., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system - 384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4
 AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamane, I., Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, F., Gissi, C., King, B., Koculwa, H., Kuenli, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quenell, P., Schriml, L.M., Staudli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Guelinich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombarts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schobach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wysshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsuke, S. and Hayashizaki, Y.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409 (6821), 685-690 (2001)
 MEDLINE 21085660
 PUBMED 11217851
 REFERENCE 5 (bases 1 to 1522)
 AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akinura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hironaka, K., Hirooka, T., Hori, F., Hume, D., Inocenti, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toya, T., Yamamuro, T., Yamane, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Substitution
 JOURNAL Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp,

/clone="TA25042G03R"
 /clone_1lb="TA25"
 /cisaue_type="developing seeds"
 /dev_stage="25 days after anthesis"
 /lab_host="B. coli DH10B"
 /note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
 Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
 of cultivar Glenlea 25 days post-anthesis"
 BASE COUNT 270 a 111 c 148 g 168 t
 ORIGIN
 Query Match 68.5%; Score 22.6; DB 14; Length 697;
 Best Local Similarity 86.2%; Pred. No. 7.8e+03;
 Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 CTGTGTTACTTCTCTCTATTCAG 30
 Db 346 CTGTGTTCTTCTCTCTCTCTAG 318
 RESULT 8 1527 bp mRNA linear EST 20-OCT-2000
 BE876684/c 601488248P1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3890914 5',
 DEFINITION mRNA sequence.
 ACCESSION BE876684
 VERSION BE876684.1 GI:10325460
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euteleostomi; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1527)
 /note="http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: DCTD/DTF/Gardar
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM674 row: 0 column: 11
 High quality sequence stop: 3.
 Location/Qualifiers
 1. 1527
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3890914"
 /clone_1lb="NIH_MGC_69"
 /cisaue_type="large cell carcinoma, undifferentiated"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.1 kb. Library constructed by Life
 Technologies."
 BASE COUNT 800 a 393 c 324 g 10 t
 ORIGIN
 Query Match 68.5%; Score 22.6; DB 12; Length 1577;
 Best Local Similarity 86.2%; Pred. No. 6.2e+03;
 Matches 25; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 CTGTGTTACTTCTCTCTATTCAG 30
 Db 1206 CTGTGTTGCTTCTCTCTCTCTCTG 1178
 RESULT 9
 BH659639

LOCUS BH659639 693 bp DNA linear GSS 19-FEB-2002
 DEFINITION BOME9191F BO_2_3_KB Brassica oleracea genomic clone BOME9191, DNA
 sequence.
 ACCESSION BH659639
 VERSION BH659639.1 GI:18718200
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 693)
 /note="Town, C.D., Van Aken, S., Utterback, T. and Frazer, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 Contact: Chris Town
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TP
 Class: sheared ends.
 Location/Qualifiers
 1. 693
 /organism="Brassica oleracea"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOME9191"
 /clone_1lb="BO_2_3_KB"
 /note="Vector: pBOS1; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pBOS1 using BstXI linkers"
 BASE COUNT 218 a 64 c 110 g 301 t
 ORIGIN
 Query Match 67.9%; Score 22.4; DB 17; Length 693;
 Best Local Similarity 81.2%; Pred. No. 8.8e+03;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 2 CTGTGTTACTTCTCTCTATTCAGCA 33
 Db 146 CTGTGTTAGTTCTCTCTATACAAACA 177
 RESULT 10
 AQ872730/c 455 bp DNA linear GSS 08-NOV-1999
 LOCUS AQ872730
 DEFINITION V15F1 mTh-3xHA/lacZ insertion library, strain AB972 Saccharomyces
 cerevisiae genomic 5', DNA sequence.
 ACCESSION AQ872730
 VERSION AQ872730.1 GI:6284876
 KEYWORDS GSS.
 SOURCE Baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 1 (bases 1 to 455)
 /note="Roes-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
 desBerges, S.A., Cheung, K.-H., Sheehan, A., Symonakis, D., Jansen, R.,
 Unanue, L., Heidman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
 Large-scale analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 Unpublished (1999)
 Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 te of mTh-3xHA/lacZ insertion.

Seq primer: GGCCCTCTTCTTGAAGTAC
 Class: transposon-tagged.
 Location/Qualifiers

1..455

/organism="Saccharomyces cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mtn-3xHA/lacZ Insertion Library, strain AB972"
 /lab_host="E. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library was prepared in pHS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ multitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 159 a 104 c 72 g 120 t

Query Match 67.3%; Score 22.2; DB 17; Length 455;
 Best Local Similarity 88.9%; Pred. No. 1.1e+04;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCTGTGTACTTCTTCTTCTATTC 27
 |||||
 Db 373 GCTGTGTGCTTCTTCTTCTGTTTC 347

RESULT 11

AO874006 473 bp DNA linear GSS 08-NOV-1999
 LOCUS V6G9 mtn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces
 DEFINITION cerevisiae genomic 5', DNA sequence.

ACCESSION AO874006
 VERSION AO874006.1 GI:6286250

KEYWORDS GSS.
 SOURCE Baker's yeast.
 ORGANISM Saccharomyces cerevisiae

Saccharomycetes; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 473)

Authors: Roess-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., deBerges, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Unanueky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
 Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption

JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161

Email: anuj.kumar@yale.edu
 te of mtn-3xHA/lacZ insertion.
 Seq primer: GGCCCTCTTCTTGAAGTAC
 Class: transposon-tagged.

FEATURES
 source Location/Qualifiers

1..473
 /organism="Saccharomyces cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mtn-3xHA/lacZ Insertion Library, strain AB972"
 /lab_host="E. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library was prepared in pHS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ multitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 155 a 108 c 84 g 125 t 1 others

ORIGIN

Query Match 67.3%; Score 22.2; DB 17; Length 473;
 Best Local Similarity 88.9%; Pred. No. 1.1e+04;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCTGTGTACTTCTTCTTCTATTC 27
 |||||
 Db 371 GCTGTGTGCTTCTTCTTCTGTTTC 345

RESULT 12

AO872889 577 bp DNA linear GSS 08-NOV-1999
 LOCUS V54C5 mtn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces
 DEFINITION cerevisiae genomic 5', DNA sequence.

ACCESSION AO872889
 VERSION AO872889.1 GI:6286133

KEYWORDS GSS.
 SOURCE Baker's yeast.
 ORGANISM Saccharomyces cerevisiae

Saccharomycetes; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 1 (bases 1 to 577)

Authors: Roess-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., deBerges, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Unanueky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
 Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption

JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161

Email: anuj.kumar@yale.edu
 te of mtn-3xHA/lacZ insertion.
 Seq primer: GGCCCTCTTCTTGAAGTAC
 Class: transposon-tagged.

FEATURES
 source Location/Qualifiers

1..577
 /organism="Saccharomyces cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mtn-3xHA/lacZ Insertion Library, strain AB972"
 /lab_host="E. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library was prepared in pHS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mtn-3xHA/lacZ multitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 180 a 147 c 105 g 145 t

ORIGIN

Query Match 67.3%; Score 22.2; DB 17; Length 577;
 Best Local Similarity 88.9%; Pred. No. 1e+04;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCTGTGTACTTCTTCTTCTATTC 27
 |||||
 Db 325 GCTGTGTGCTTCTTCTTCTGTTTC 299

RESULT 13

AO872952 578 bp DNA linear GSS 08-NOV-1999
 LOCUS V56B3 mtn-3xHA/lacZ Insertion Library, strain AB972 Saccharomyces
 DEFINITION cerevisiae genomic 5', DNA sequence.

ACCESSION AO872952
 VERSION AO872952.1 GI:6285196

KEYWORDS GSS.
 SOURCE Baker's yeast.

ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE 1 (bases 1 to 578)
 AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A.,
 deBraga, S. A., Cheung, K.-H., Sheehan, A., Symoniatie, D., Jansen, R.,
 Umanaky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G. S. and Snyder, M.
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumaryale.yale.edu
 te of mtH-3xHA/lacZ insertion.
 Seq primer: GGCTCTCTTCTTCTTGAAGTAC
 Class: transposon-tagged.
 FEATURES
 source location/Qualifiers
 1..578
 /organism="Saccharomyces cerevisiae"
 /strain="AB972 - trp1 r(0) (S288C background)"
 /db_xref="taxon:4932"
 /clone_lib="mtH-3xHA/lacZ Insertion library, strain AB972"
 /lab_host="E. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library was
 prepared in pHS6-Sal; genomic DNA was size-fractionated
 (DNA of roughly 2-3 kb in length) prior to cloning. This
 library was subsequently mutagenized with a mtH-3xHA/lacZ
 multitransposon containing lacZ, URA3, and tet resistance."
 BASE COUNT 185 a 143 c 101 g 149 t
 ORIGIN
 Query Match 67.3%; Score 22.2; DB 17; Length 578;
 Best Local Similarity 88.9%; Pred. No. 1e+04; Indels 0; Gaps 0;
 Matches 24; Conservative 0; Mismatches 3;
 QY 1 GCTGTGTTACTTTCTTCTTCTATTC 27
 |||||
 Db 362 GCTGTGTTGCTTTCTTCTGTTCTGTC 336
 RESULT 14
 BG850976 708 bp mRNA linear EST 29-MAY-2001
 LOCUS 1024029H08.Y1 C. reinhardtii CC-1690, normalized, Lambda Zap II
 DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION BG850976
 VERSION BG850976.1 GI:14232160
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadales; Chlamydomonadae.
 REFERENCE 1 (bases 1 to 708)
 AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P.,
 McDermott, J. P., Silflow, C., Stern, D., and Surzycki, R.
 TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants; project phase 2
 JOURNAL Unpublished (2000)
 COMMENT Contact: Charles Hauser
 DCMB Box 91000
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu.
 FEATURES
 location/Qualifiers

source
 1..708
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
 II"
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
 XhoI; This library, constructed by John Davies and Jeffrey
 McDermott, combines cDNAs from CC-1690 cells grown to
 mid-log phase in TAP (acetate-containing) medium in the
 light, TAP medium in the dark, HS (minimal) medium in
 ambient levels of CO2 and HS medium bubbled with 5% CO2.
 PolyA mRNA was purified from each sample, pooled and cDNA
 synthesized. The cDNA was directionally cloned into lambda
 Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
 pBluescript II SK- plasmids were excised from the lambda
 Zap clones by superinfection with ExAssist (Stratagene)
 phage. The library was normalized using method 4 described
 in Bonaldo et al (1996) Genome Research 6: 791-806."
 BASE COUNT 93 a 134 c 105 g 376 t
 ORIGIN
 Query Match 67.3%; Score 22.2; DB 12; Length 708;
 Best Local Similarity 88.9%; Pred. No. 9.8e+03; Indels 0; Gaps 0;
 Matches 24; Conservative 0; Mismatches 3;
 QY 2 CTGTGTTACTTTCTTCTTCTATTC 28
 |||||
 Db 435 CTGTGTTATTTTCTTCTTCTATTC 461
 RESULT 15
 AQ876136 810 bp DNA linear GSS 08-NOV-1999
 LOCUS V133f7 mtH-3xHA/lacZ Insertion library, strain Y2278 Saccharomyces
 cerevisiae genomic 5', DNA sequence.
 ACCESSION AQ876136
 VERSION AQ876136.1 GI:6288380
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE 1 (bases 1 to 810)
 AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P. S. R., Agarwal, S., Kumar, A.,
 deBraga, S. A., Cheung, K.-H., Sheehan, A., Symoniatie, D., Jansen, R.,
 Umanaky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
 Hager, K., Miller, P., Roeder, G. S. and Snyder, M.
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumaryale.yale.edu
 te of mtH-3xHA/lacZ insertion.
 Seq primer: GGCTCTCTTCTTCTTGAAGTAC
 Class: transposon-tagged.
 FEATURES
 source location/Qualifiers
 1..810
 /organism="Saccharomyces cerevisiae"
 /strain="Y2278 - S288C background, c1r(0) rho(0)"
 /db_xref="taxon:4932"
 /clone_lib="mtH-3xHA/lacZ Insertion library, strain Y2278"
 /lab_host="E. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library
 without 2 micron or mitochondrial DNA was prepared in
 pHS6-Sal; genomic DNA was size-fractionated (DNA of
 roughly 2-3 kb in length) prior to cloning. This library
 was subsequently mutagenized with a mtH-3xHA/lacZ

"Minttransposon containing lacZ, URA3, and tet resistance.
 BASE COUNT 222 a 158 c 168 g 260 t 2 others
 ORIGIN

Query Match 67.3%; Score 22.2; DB 17; Length 810;
 Best Local Similarity 88.9%; Pred. No. 9.4e+03;
 Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCGGTGTACTTCTCTCTATTC 27
 |||||||
 Db 173 GCGGTGTCTCTCTCTCTCTGTC 199

Search completed: May 11, 2003, 04:54:30
 Job time : 535.481 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 60.2907 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-17
Perfect score: 31
Sequence: 1 attccacccttaccctccgccttcattg 31

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	21	AAA96851
2	20.4	65.8	339	14	AA061194
3	20.4	65.8	1288	20	AAK51414
4	20.4	65.8	40962	23	ABL29926
5	20.4	65.8	2944528	24	ABA03041
6	19.8	63.9	498	24	ABL6989
7	19.8	63.9	3079	22	AAH4182
8	19.4	62.6	291	24	ABK76620
9	19.4	62.6	595	22	ABA61610

C 10	19.4	62.6	595	22	AA141521	Probe #10207 used
C 11	19.4	62.6	1671	23	AA553289	Haemophilus influe
C 12	19.4	62.6	1630121	17	AAT42063	Haemophilus influe
C 13	19.2	61.9	537	22	AA914400	N. meningitidis (s
C 14	19.2	61.9	537	24	ABK37780	Strong promoter fr
C 15	19.2	61.9	1000	22	AA913199	DNA sequence upstr
C 16	19.2	61.9	1000	24	ABK37779	DNA sequence upstr
C 17	19.2	61.9	1260	23	AA594538	DNA encoding novel
C 18	19.2	61.9	1281	9	AA81257	Mutant ara gene e
C 19	19.2	61.9	1281	13	AA029507	EPSP synthase wild
C 20	19.2	61.9	1281	13	AA029508	EPSP synthase muta
C 21	19.2	61.9	1281	15	AA064903	EPSP synthase gene
C 22	19.2	61.9	1281	15	AA064904	Mutant EPSP syntha
C 23	19.2	61.9	1348	22	AA06954	ara gene encoding
C 24	19.2	61.9	2056	23	AA593259	DNA encoding novel
C 25	19.2	61.9	2462	23	AA593184	DNA encoding novel
C 26	19.2	61.9	2562	23	AA573488	DNA encoding novel
C 27	19.2	61.9	4530	22	AA06958	pme4 complementing ar
C 28	19.2	61.9	30078	21	AA081520	N. meningitidis pa
C 29	19.2	61.9	49767	21	AA081458	N. meningitidis pa
C 30	19.2	61.9	56485	21	AA081476	N. meningitidis pa
C 31	19.2	61.9	172325	21	AA081476	N. meningitidis pa
C 32	19.2	61.9	349980	21	AA081476	N. meningitidis pa
C 33	19.2	61.9	349980	21	AA081476	N. meningitidis pa
C 34	19.2	61.9	837096	21	AA081489	N. meningitidis pa
C 35	19.2	61.9	1437668	21	AA081490	N. meningitidis B
C 36	19.2	61.9	735	21	AA081490	N. meningitidis B
C 37	19.2	61.3	811	21	AA049886	Arabidopsis thalia
C 38	19.2	61.3	812	21	AA040364	Arabidopsis thalia
C 39	19.2	61.3	1033	23	AA089565	DNA encoding novel
C 40	19.2	61.3	1108	23	AA089561	DNA encoding novel
C 41	19.2	61.3	1362	21	AA047934	Arabidopsis thalia
C 42	19.2	61.3	14568	24	AB132230	Human immune syste
C 43	19.2	61.3	23683	24	AB170482	Chemically treated
C 44	19.2	61.3	23683	24	AB170482	Human metacelsis a
C 45	18.8	60.6	319	22	ABA11402	Human nervous syst

ALIGNMENTS

RESULT 1	AAA96851	standard; DNA; 31 BP.
ID	AAA96851	
AC	AAA96851:	
XX		
DT	19-FEB-2001	(first entry)
XX		
DE	Guide desoxynucleotide building block G3.	
XX		
KM	Promoter, intergenic region; Commelina yellow mottle virus;	
KW	chimeric expression promoter; plant vascular expression promoter;	
KM	plant green tissue expression promoter; Cassava vein mosaic virus;	
KW	transgenic plant; ss.	
XX		
OS	Synthetic.	
XX		
PN	WO200058485-A1.	
XX		
PD	05-OCT-2000.	
XX		
PF	29-MAR-2000; 2000WO-IB00370.	
XX		
PR	29-MAR-1999; 99FR-0003925.	
XX		
PA	(MERI-) MERISTEM THERAPEUTICS.	
XX		
PI	Rance I, Gruber V, Theisen M;	
XX		
DR	WPI; 2000-647238/62.	
XX		
PT	Chimeric expression promoter for transgenic plant production, comprises	

PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Disclosure; Page 24, 91pp; English.
XX
CC The present sequence represents a guide desoxynucleotide building
CC block, which was used to link directional desoxynucleotide building
CC blocks during construction of chimeric expression promoters. These
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comelina yellow mottle virus, and the second plant promoter originates
CC from the Casava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SO Sequence 31 BP; 5 A; 11 C; 2 G; 13 T; 0 other;
XX
Query Match 100.0%; Score 31; DB 21; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.0007;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ATTCACCTTTTACATTCGCCGCTTCAATG 31
Db 1 ATTCACCTTTTACATTCGCCGCTTCAATG 31
XX
RESULT 2
ID AAG61194/c
AC AAG61194; standard; DNA; 339 BP.
XX
XX AAG61194;
XX
DT 16-MAR-1994 (first entry)
XX
DE Human brain Expressed Sequence Tag EST01232.
XX
XX Gene transcription product; genetic markers; tagging; in vivo;
XX transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
OS Homo sapiens.
XX
XX WO9316178-A.
XX
XX 19-AUG-1993.
XX
XX 12-FEB-1993; 93WO-US01234.
XX
XX 12-FEB-1992; 92US-0837195.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX
XX Adams MD, Moreno RF, Venter CJ;
XX
XX WPI; 1993-272882/34.
XX
XX Enriched oligonucleotides and corresp. sequences - used as
XX PT markers for human genes transcribed in-vivo, facilitate tagging
XX PT of most human genes
XX
XX Example 4; Page 452; 500pp; English.
XX
XX The Expressed Sequence Tag was isolated from a human brain cDNA
XX library as part of a large set of ESTs which can be used as markers
XX for human genes transcribed in vivo. They can be used to facilitate
XX tagging of most human genes, for mapping locations of expressed genes
XX on chromosomes, for individual or forensic identification, for mapping
XX locations of disease-associated genes, for identification of tissue
XX type, and for prepu. of antisense sequences, probes and constructs.

CC EST01232 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also AAG59041-Q61440.
XX
XX
SO Sequence 339 BP; 115 A; 66 C; 66 G; 92 T; 0 other;
XX
Query Match 65.8%; Score 20.4; DB 14; Length 339;
Best Local Similarity 80.0%; Pred. No. 31;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
OY 2 TTCACCTTTTACATTCGCCGCTTCAATG 31
Db 76 TTCACCTTTTACATTCGCCGCTTCAATG 47
XX
RESULT 3
ID AAX61414
AC AAX61414; standard; cDNA; 1288 BP.
XX
XX AAX61414;
XX
DT 14-JUL-1999 (first entry)
XX
XX DNA encoding a human secreted protein.
XX
DE Human secreted protein; cancer; tumour; neurodegenerative disorder;
XX developmental abnormality; fetal deficiency; blood disorder; leukemia;
XX immune system disease; autoimmune disease; hepatic disease;
XX renal disease; lymphoma; inflammation; allergy; ischemic shock;
XX Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
XX obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
XX lung disease; thymus disease; digestive disorder; endocrine disorder;
XX infection; AIDS; ss.
XX
XX Homo sapiens.
XX
XX WO922243-A1.
XX
XX 06-MAY-1999.
XX
XX 23-OCT-1998; 98WO-US22376.
XX
XX 24-OCT-1997; 97US-0063387.
XX 24-OCT-1997; 97US-0062784.
XX 24-OCT-1997; 97US-0063088.
XX 24-OCT-1997; 97US-0063089.
XX 24-OCT-1997; 97US-0063090.
XX 24-OCT-1997; 97US-0063091.
XX 24-OCT-1997; 97US-0063092.
XX 24-OCT-1997; 97US-0063098.
XX 24-OCT-1997; 97US-0063099.
XX 24-OCT-1997; 97US-0063100.
XX 24-OCT-1997; 97US-0063101.
XX 24-OCT-1997; 97US-0063109.
XX 24-OCT-1997; 97US-0063110.
XX 24-OCT-1997; 97US-0063111.
XX 24-OCT-1997; 97US-0063148.
XX 24-OCT-1997; 97US-0063386.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
XX PI Peng P, Florence C, Florence KA, Greene JM, Janat F;
XX PI Kaye H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
XX PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
XX
XX WPI; 1999-303069/25.
XX
XX P-PSDB; AAY19534.
XX
XX New isolated human genes and the secreted polypeptides they encode
XX PT Claim 3; Page 359; 546pp; English.
XX

CC The specification describes cDNA sequences (AAK61322-X61470) encoding
 CC human secreted proteins (AAV19442-Y19590). The polynucleotides and their
 CC corresponding secreted polypeptides are useful for preventing, treating
 CC or ameliorating medical conditions, e.g. by protein or gene therapy.
 CC Pathological conditions can also be diagnosed by determining the amount
 CC of the polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the polynucleotides, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and fetal deficiencies, blood disorders, leukemias,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, prostate diseases, obesity,
 CC disorders involving osteoclasts such as osteoporosis, arthritis or
 CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
 CC disorders, infections and AIDS. The polypeptides are also useful for
 CC identifying their binding partners.

XX Sequence 1288 BP; 333 A; 304 C; 237 G; 407 T; 7 other;

Query Match 65.8%; Score 20.4; DB 20; Length 1288;
 Best Local Similarity 80.0%; Pred. No. 39;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

2 TTCACCTTTACATCCCGCTTTCATTG 31
 1000 TACCACTTTAAATCCAGCTCATCTG 1029

RESULT 4

ABL29926/c
 ID ABL29926 standard; DNA; 40962 BP.

XX ABL29926;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 41251.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li FWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Claim 1; SEQ ID NO 41251; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins

CC (AB57737-AB872072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 40962 BP; 12314 A; 8661 C; 8334 G; 11653 T; 0 other;

Query Match 65.8%; Score 20.4; DB 23; Length 40962;
 Best Local Similarity 95.5%; Pred. No. 70;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

7 CTTTACATTCCTCCGCTTCA 28
 11690 CTTTACATTCCTCCGCTTCA 11669

RESULT 5
 ABA03041/c
 ID ABA03041 standard; DNA; 2944528 BP.

XX ABA03041;

XX 05-FEB-2002 (first entry)

XX Listeria monocytogenes EGD-e genome sequence.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

XX vitamin B12; bacterial infection; disease; ds.

XX Listeria monocytogenes.

XX WO200177335-A2.

XX 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR01118.

XX 11-APR-2000; 2000FR-0004629.

XX (INSP) INST PASTER.

XX Buchrieser C, Frangoul L, Couve E, Rusniok C, Feihl H, Dehoux P,
 PI Dussurget O, Chetoui H, Nedjati H, Glaser P, Kunst F, Cossart P,
 PI Daniels J, Goebel W, Kneft J, Kuhn M, Ng E, Vazquez-Boland JA,
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L,
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J,
 PI Rose M, Voss H;

XX WPI; 2002-010914/01.

XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -

XX Claim 1; SEQ ID NO 1; 192pp; French.

XX The present sequence is the genome sequence of Listeria monocytogenes
 CC EGD-e. This sequence and fragments of this sequence are useful for
 CC selecting probes and primers for detecting genes in L. monocytogenes and
 CC related organisms, and to study genetic polymorphisms and other genomes.
 CC Proteins (ABB47297-ABB50149) expressed from the present sequence are
 CC useful for raising specific antibodies, identification of L.
 CC monocytogenes and related organisms, and for biosynthesis and
 CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
 CC proteins encoded by it are also useful for selecting compounds that
 CC regulate gene expression and cell replication and modulate L.
 CC monocytogenes-related diseases. In addition, this sequence and proteins
 CC encoded by it are useful in pharmaceutical and vaccine compositions for
 CC the treatment or prevention of infections by L. monocytogenes and related
 CC organisms.
 CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SO Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;
 Query/Match 65.8%; Score 20.4; DB 24; Length 2944528;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 TTCGACCTTTACATTCGCGCTTCATTG 31
 DB 745445 TTCGACCTTTACATTCGCGCTTCATTG 745416

RESULT 6
 ABL66989/c
 ID ABL66989 standard; DNA; 498 BP.

AC ABL66989;
 XX
 DT 15-MAY-2002 (first entry)
 XX

DE Thyroid cancer related gene sequence SEQ ID NO:5326.
 XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KM stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KM cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
 KM gene; ds.

OS Homo sapiens.
 XX
 PN WO200194629-A2.
 XX

PD 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US10838.

XX
 PR 05-JUN-2000; 2000US-209473P.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-231133P.
 PR 18-SEP-2000; 2000US-233617P.
 PR 20-SEP-2000; 2000US-234099P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 20-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234509P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234923P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235637P.
 PR 26-SEP-2000; 2000US-235638P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235840P.
 PR 27-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236033P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.

PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.

XX (AVAL-) AVALON PHARM.

PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
 PI Soppe DR, Weaver Z;

XX WPI; 2002-188264/24.

PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set -
 XX
 XX Claim 1; SEQ ID 5326; 44pp; English.

CC The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer.
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilm's tumour.
 XX

SO Sequence 498 BP; 126 A; 85 C; 121 G; 164 T; 2 other;

Query Match 63.9%; Score 19.8; DB 24; Length 498;
 Best Local Similarity 77.4%; Pred. No. 60;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 ATTCGACCTTTACATTCGCGCTTCATTG 31
 DB 158 ATTCGACCTTTTAAAGTCCACCTTTCATTG 128

RESULT 7

AAH14182
 ID AAH14182 standard; cDNA; 3079 BP.

XX AAH14182;
 AC
 XX 26-JUN-2001 (first entry)
 DT
 XX

DE Human cDNA sequence SEQ ID NO:11429.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 KM
 XX Homo sapiens.
 OS

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.

XX	PR	09-JUN-2000; 2000JP-0241899.
XX	PA	(HELI-) HELIX RES INST.
XX	P1	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX	P1	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	XX	WPI, 2001-318749/34.
XX	PT	primer sets for synthesizing polynucleotides, particularly the 5602
XX	PT	full-length cDNAs defined in the specification, and for the detection
XX	PT	and/or diagnosis of the abnormality of the proteins encoded by the
XX	PT	full-length cDNAs -
PS	Claim 8,	SEQ ID 11429; 2537BP + CD ROM; English.
XX	XX	The present invention describes primer sets for synthesizing 5602
CC	CC	full-length cDNAs defined in the specification, where a primer set
CC	CC	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC	CC	to the complementary strand of a polynucleotide which comprises one of
CC	CC	the 5602 nucleotide sequences defined in the specification, where the
CC	CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC	CC	of an oligonucleotide comprising a sequence complementary to the
CC	CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	CC	polynucleotide which comprises a 3'-end sequence, where the
CC	CC	oligonucleotide comprises at least 15 nucleotides and the combination of
CC	CC	the 5'-end sequence/3'-end sequence is selected from those defined in
CC	CC	the specification. The primer sets can be used in antisense therapy and
CC	CC	in gene therapy. The primers are useful for synthesizing polynucleotides,
CC	CC	particularly full-length cDNAs. The primers are also useful for the
CC	CC	detection and/or diagnosis of the abnormality of the proteins encoded by
CC	CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	CC	cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
CC	CC	AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC	CC	AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC	CC	represent oligonucleotides, all of which are used in the exemplification
CC	CC	of the present invention.
SQ	Sequence	3079 BP; 890 A; 664 C; 575 G; 950 T; 0 other;
Query Match	63.9%;	Score 19.8; DB 22; Length 3079;
Best Local Similarity	77.4%;	Fred. NO. 81;
Matches 24; Conservative	0;	Mismatches 7; Indels 0; Gaps 0;
OY	1	ATTCACCGTTTACACTGCCGCTTCATTG 31
DB	1217	ATAACAATTTTAAAGTCCCACCCTTTCATG 1247
RESULT 8		
ID	ABK76620	standard; DNA; 291 BP.
AC	ABK76620;	
XX	13-AUG-2002	(first entry)
DT	Bacillus licheniformis genomic sequence tag (GST) #3911.	
DE	Differential gene expression; genomic sequenced tag; GST;	
KW	altered culture condition; environmental stress;	
KW	physiological provocation; ds.	
OS	Bacillus licheniformis.	
XX	WO200229113-A2.	
PN	11-APR-2002.	
PD	05-OCT-2001; 2001WO-US31437.	
XX	06-OCT-2000; 2000US-0680598.	
PR		

PR 27-MAR-2001; 2001US-279526P.
XX (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
PI Berka R, Clausen IG;
XX
DR WPI; 2002-416684/44.
XX
PT Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -
XX
PS Claim 4; SEQ ID NO 3911; 200pp; English.
XX
XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 291 BP; 96 A; 55 C; 39 G; 101 T; 0 other;
XX
Query Match 62.6%; Score 19.4; DB 24; Length 291;
Best Local Similarity 79.3%; Pred. No. 81;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
CY 2 TTCCACCTTTACATTCGCGCTTTCAAT 30
DB 256 TTCCCTTATTCATACCCCTTTCAAT 284
RESULT 9
ABA61610/c
ID ABA61610 standard; DNA; 595 BP.
XX
XX ABA61610;
XX
DT 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #9915.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
PR

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
DR
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 9915; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 595 BP; 194 A; 89 C; 110 G; 202 T; 0 other;
Query Match 62.6%; Score 19.4; DB 22; Length 595;
Best Local Similarity 79.3%; Pred. No. 91;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
OY 2 TTCACCTTTACATTCGCGCTTCATT 30
Db 64 TTCACCTTCACCTTCGCTTCATT 36
RESULT 10
AA141521/c
ID AA141521 standard; DNA; 595 BP.
XX
AC AA141521;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #10207 used to measure gene expression in human placenta sample.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 10207; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 595 BP; 194 A; 89 C; 110 G; 202 T; 0 other;
Query Match 62.6%; Score 19.4; DB 22; Length 595;
Best Local Similarity 79.3%; Pred. No. 91;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
OY 2 TTCACCTTTACATTCGCGCTTCATT 30
Db 64 TTCACCTTCACCTTCGCTTCATT 36
RESULT 11
AAS53289/c
ID AAS53289 standard; DNA; 1671 BP.
XX
AC AAS53289;
XX
DT 13-FEB-2002 (first entry)
XX
DE Haemophilus influenzae DNA for cellular proliferation protein #71.
XX
XX Antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX
XX Haemophilus influenzae.
XX
XX WO200170955-A2.
XX
PD 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
PF 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207272P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR P-PDB; AAU35430.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 6926; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programs. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1671 BP, 447 A, 276 C, 419 G, 529 T, 0 other;
Query Match 62.6%; Score 19.4; DB 23; Length 1671;
Best Local Similarity 79.3%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TTCCACCTTTTACATTCGCCGCTTTCATT 30
DB 421 TTTCACCTTTTACATTCACACCTTGCTTT 393
RESULT 12
AAAT42063
ID AAT42063 standard; DNA, 1830121 BP.
XX
AC AAT42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KM Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN WO9633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX
PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UWJO) UNIV JOHNS HOPKINS.
XX
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
DR WPI; 1996-485782/48.
XX
PT Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (1), a representative fragment of (1) or a nucleotide
CC sequence at least 9% identical to (1). By providing the full-length
CC genome sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
XX

SQ Sequence 1830121 BP, 567399 A, 350615 C, 347389 G, 564036 T, 682 other;
Query Match 62.6%; Score 19.4; DB 17; Length 1830121;
Best Local Similarity 79.3%; Pred. No. 3e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 2 TTCCACCTTTTACATTCGCCGCTTTCATT 30
DB 468917 TTTCACCTTTTACATTCACACCTTGCTTT 468945
RESULT 13
AAAF91400/C
ID AAF91400 standard; DNA, 537 BP.
XX
AC AAF91400;
XX
DT 04-MAY-2001 (first entry)
XX
DE N. meningitidis (serogroup B) PorB promoter region, SEQ ID:26.
XX
KM Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW genetically modified; protective antigen expression; LPS detoxification;
KW LPS; lipid A; homologous recombination vector; immunisation;
KW immunoprotective; non-toxic; paediatric; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200109350-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07424.
XX
PR 03-AUG-1999; 99GB-0018319.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthel FJ, Dalemans WJ, Denoel P, Deguesne G, Feron C, Lobet Y;
PI Poolman J, Thiry G, Thonard J, Voet P;
XX
DR WPI; 2001-138654/14.
XX
PT New isolated polynucleotide useful for outer membrane vesicle
PT preparation from Gram-negative bacterial strain for vaccination of
PT microbial infections -
XX
PS Disclosure; Page 83; 128pp; English.
XX
CC The invention relates to a genetically-engineered outer membrane vesicle
CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC The blebs of the invention are improved with respect to their
CC immunogenicity and toxicity by the introduction of one or more genetic
CC changes to the chromosome of the bacterium from which the blebs are
CC derived. The changes made include the upregulation of protective antigen
CC expression, the downregulation of immunodominant non-protective antigen
CC expression, and genetic changes which result in detoxification of the
CC Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC encompasses modified Gram-negative bacterial strains from which the bleb
CC preparations are made, a vector suitable for performing recombination
CC events (for the generation of the modified bacterial strains),
CC bacterially-derived nucleic acid sequences used in such a vector, and an
CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC cell vaccine suitable for paediatric use. The bleb preparation is useful
CC in the manufacture of a medicament for immunising a human host against a
CC disease caused by infection of one or more of the following: Neisseria
CC meningitidis, Neisseria gonorrhoeae, Haemophilus influenza, Moraxella
CC catarrhalis, Pseudomonas aeruginosa, Chlamydia trachomatis, and Chlamydia
CC pneumoniae. The invention may also be used to provide immunisation against
CC the influenza virus. Bacterially derived nucleotide sequences of the
CC invention are used in the performance of homologous recombination events
CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC increase or decrease expression of that gene. Immunoprotective and

CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC are more immunogenic, less toxic and safer, and are particularly useful
CC for paediatric use. The present sequence represents a *Neisseria*
CC meningitidis promoter sequence.

XX
SQ Sequence 537 BP, 124 A; 102 C; 153 G; 158 T; 0 other;

Query Match 61.9%; Score 19.2; DB 22; Length 537;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATTCACCTTTACATCCCGCCT 24
DB 114 ATTCACGTTCTAGATCCCGCCT 91

RESULT 14
ABK37780/c
ID ABK37780 standard; DNA; 537 BP.
XX
AC ABK37780;
XX
DT 08-MAY-2002 (first entry)
XX
DE Strong promoter from PorB gene.
XX
KM Promot; ds; Antibacterial; vaccine; bleb; Gram-negative bacteria;
KM outer membrane; LPS; lipopolysaccharide; meningitis; bacteraemia;
KM otitis media; pneumonia; chronic bronchitis; sinusitis.
XX
OS *Neisseria meningitidis* serogroup B.
XX
PN WO200209746-A2.
XX
PD 07-FEB-2002.
XX
PF 31-JUL-2001; 2001WO-EP08857.
XX
PR 31-JUL-2000; 2000EP-0956369.
PR 08-FEB-2001; 2001GB-0003170.
XX
PA (SMIK) *SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthet FJ, Dalemans W, Denoel P, Dequeane G, Feron C, Garcon N;
PI Lobet Y, Poolman J, Thiry G, Thonnard J, Voet P;
XX
DR WPI; 2002-188688/24.
XX
PT New immunogenic composition comprising an antigen derived from a
PT pathogen and a bleb preparation from *Neisseria meningitidis*, useful as
PT a vaccine for treating or preventing disease caused by the pathogen -
XX
PS Disclosure; Page 86; 125pp; English.

CC The invention relates to an immunogenic composition comprising an antigen
CC derived from a pathogen capable of protecting a host against the
CC pathogen, mixed with an adjuvant comprising a bleb preparation derived
CC from a Gram-negative bacterial strain. The immunogenic composition
CC consists of *N. meningitidis* B blebs or *N. meningitidis* C polysaccharide
CC antigen. The blebs (derived from the outer membrane) may also
CC have their toxic lipopolysaccharide (LPS) content reduced using
CC heterologous down regulating sequences for LPS pathway genes or by
CC up regulating genes involved in LPS synthesis suppression, by
CC a promoter replacement technique. The immunogenic preparation is useful
CC in the manufacture of a medicament for the treatment of a disease caused
CC by the pathogen from which the antigen is derived (e.g. from
CC *Neisseria*, meningitis and bacteraemia, from *Moraxella*, otitis media and
CC pneumonia, and from *H. influenzae* chronic bronchitis, sinusitis,
CC pneumonia and otitis media). The bleb derived from *M. catarrhalis* or
CC from a non-typeable *H. influenzae* is useful as an adjuvant in an
CC immunogenic composition comprising one or more pneumococcal capsular
CC polysaccharides or protein antigens. The present sequence is a strong
CC promoter suitable for constitutively expressing genes in

CC *N. meningitidis*.
XX
SQ Sequence 537 BP, 124 A; 102 C; 153 G; 158 T; 0 other;

Query Match 61.9%; Score 19.2; DB 24; Length 537;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATTCACCTTTACATCCCGCCT 24
DB 114 ATTCACGTTCTAGATCCCGCCT 91

RESULT 15
AAF91399/c
ID AAF91399 standard; DNA; 1000 BP.
XX
AC AAF91399;
XX
DT 04-MAY-2001 (first entry)
XX
DE *N. meningitidis* (serogroup A) PorB gene upstream sequence, SEQ ID:25.
XX
KM Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KM genetically modified; protective antigen expression; LPS detoxification;
KM LPS; lipid A; homologous recombination vector; immunisation;
KM immunoprotective; non-toxic; paediatric; cyclic; circular; ds.
XX
OS *Neisseria meningitidis*.
XX
PN WO200109350-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07424.
XX
PR 03-AUG-1999; 99GB-0018319.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Berthet FJ, Dalemans WJ, Denoel P, Dequeane G, Feron C, Lobet Y;
PI Poolman J, Thiry G, Thonnard J, Voet P;
XX
DR WPI; 2001-138654/14.
XX
PT New isolated polynucleotide useful for outer membrane vesicle
PT preparation from Gram-negative bacterial strain for vaccination of
PT microbial infections -
XX
PS Claim 46; Page 83; 128pp; English.

CC The invention relates to a genetically-engineered outer membrane vesicle
CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
CC The blebs of the invention are improved with respect to their
CC immunogenicity and toxicity by the introduction of one or more genetic
CC changes to the chromosome of the bacterium from which the blebs are
CC derived. The changes made include the upregulation of protective antigen
CC expression, the downregulation of immunodominant non-protective antigen
CC expression, and genetic changes which result in detoxification of the
CC lipid A moiety of lipopolysaccharide (LPS). The invention also
CC encompasses modified Gram-negative bacterial strains from which the bleb
CC preparations are made, a vector suitable for performing recombination
CC events for the generation of the modified bacterial strains).
CC bacterially-derived nucleic acid sequences used in such a vector, and an
CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC cell vaccine suitable for paediatric use. The bleb preparation is useful
CC in the manufacture of a medicament for immunising a human host against a
CC disease caused by infection of one or more of the following: *Neisseria*
CC meningitidis, *Neisseria gonorrhoeae*, *Haemophilus influenzae*, *Moraxella*
CC catarrhalis, *Pseudomonas aeruginosa*, *Chlamydia trachomatis*, and *Chlamydia*
CC pneumonia. The invention may also be used to provide immunisation against
CC the influenza virus. Bacterially derived nucleotide sequences of the
CC invention are used in the performance of homologous recombination events

CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
 CC increase or decrease expression of that gene. Immunoprotective and
 CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
 CC are more immunogenic, less toxic and safer, and are particularly useful
 CC for paediatric use. The present sequence represents a specifically
 CC claimed *Neisseria meningitidis* nucleic acid sequence.

XX Sequence 1000 BP; 264 A; 223 C; 240 G; 273 T; 0 other;

Query Match 61.9%; Score 19.2; DB 22; Length 1000;
 Best Local Similarity 87.5%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATTCACCTTTACATTCGCCCT 24
 |||||
 Db 114 ATTCACCTTCTAGATTCCGCCCT 91

Search completed: May 11, 2003, 03:16:48
 Job time : 658.291 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 12.4311 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-17

Perfect score: 31
Sequence: 1 attccacatttaccatccgccttcattg 31

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents NA.*
2: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/BACKFILE1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.8	63.9	856	4	US-08-998-416-308 Sequence 308, App
2	19.2	61.9	1348	4	US-09-064-693A-22 Sequence 22, Appl
3	19.2	61.9	4530	4	US-09-064-693A-26 Sequence 26, Appl
4	18.8	60.6	6693	4	US-08-961-527-195 Sequence 195, App
5	18.4	59.4	2465	3	US-08-619-812-5 Sequence 5, Appl
6	18.2	58.7	7070	1	US-08-619-554-3 Sequence 3, Appl
7	18	58.1	298	4	US-09-071-035-147 Sequence 147, App
8	18	58.1	441	4	US-09-071-035-145 Sequence 145, App
9	18	58.1	502	3	US-08-961-083-137 Sequence 137, App
10	18	58.1	25002	4	US-08-961-527-48 Sequence 48, Appl
11	17.8	57.4	685	4	US-09-328-111-101 Sequence 101, App
12	17.6	56.8	2765	4	US-09-397-885-2 Sequence 2, Appl
13	17.4	56.1	1202	3	US-09-058-489-29 Sequence 29, Appl
14	17.4	56.1	5300	4	US-08-938-669A-2 Sequence 1, Appl
15	17.4	56.1	5304	4	US-08-938-669A-2 Sequence 2, Appl
16	17.4	56.1	6169	4	US-08-938-669A-3 Sequence 3, Appl
17	17.4	55.5	658	4	US-09-470-191-57 Sequence 57, Appl
18	17.2	55.5	1485	4	US-08-976-259-13 Sequence 13, Appl
19	17.2	55.5	2337	4	US-09-134-001C-1031 Sequence 1031, Ap
20	17.2	55.5	3390	1	US-08-453-742-26 Sequence 26, Appl
21	17.2	55.5	3390	1	US-08-453-742-26 Sequence 26, Appl
22	17.2	55.5	3390	1	US-08-453-742-26 Sequence 26, Appl
23	17.2	55.5	3390	1	US-08-453-742-26 Sequence 26, Appl
24	17.2	55.5	3416	1	US-08-453-742-24 Sequence 24, Appl
25	17.2	55.5	3416	1	US-08-453-742-24 Sequence 24, Appl
26	17.2	55.5	3416	1	US-08-453-742-24 Sequence 24, Appl
27	17.2	55.5	3416	1	US-08-453-742-24 Sequence 24, Appl

C	28	17.2	55.5	3588	1	US-07-792-885A-2	Sequence 2, Appl
	29	17.2	55.5	49795	4	US-09-453-702B-60	Sequence 60, Appl
C	30	17	54.8	522	2	US-08-909-965C-16	Sequence 16, Appl
	31	17	54.8	755	4	US-09-020-955C-21	Sequence 21, Appl
	32	17	54.8	755	4	US-09-030-607-21	Sequence 21, Appl
	33	17	54.8	755	4	US-09-605-785-21	Sequence 21, Appl
	34	17	54.8	755	4	US-09-439-313-21	Sequence 21, Appl
	35	17	54.8	755	4	US-09-352-616A-21	Sequence 21, Appl
	36	17	54.8	755	4	US-09-332-149A-21	Sequence 21, Appl
	37	17	54.8	2308	4	US-09-377-155-1	Sequence 1, Appl
	38	17	54.8	2308	4	US-09-669-974-1	Sequence 1, Appl
C	39	17	54.8	3747	2	US-09-080-897-1	Sequence 1, Appl
C	40	17	54.8	3747	4	US-09-323-735-1	Sequence 1, Appl
C	41	17	54.8	5822	3	US-08-899-595-4	Sequence 4, Appl
	42	17	54.8	5822	3	US-08-899-595-5	Sequence 5, Appl
	43	17	54.8	45546	4	US-09-146-053-6	Sequence 6, Appl
	44	16.8	54.2	1252	4	US-09-276-531-92	Sequence 92, Appl
C	45	16.8	54.2	1987	4	US-09-227-357-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-08-998-416-308
Sequence 308, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Rebschnung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Weigis, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 308:
SEQUENCE CHARACTERISTICS:
LENGTH: 856 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1251UP

US-08-998-416-308

Query Match 63.9%; Score 19.8; DB 4; Length 856;
Best Local Similarity 77.4%; Pred. No. 10;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATTCACCTTTTACATCCCGCTTCAATG 31
|||
Db 293 ATTACCCCATTAATTCGCGCTTTATG 323

RESULT 2

US-09-064-693A-22/C
; Sequence 22, Application US/09064693A
; Patent No. 6210937

; GENERAL INFORMATION:

; APPLICANT: Ward, Thomas E.

; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY

; TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: W. Gary Goodson

; ADDRESSEE: INEL--Lockheed Martin Idaho

; ADDRESSEE: Technologies Co.

; STREET: P.O. Box 1625

; CITY: Idaho Falls

; STATE: Idaho

; COUNTRY: USA

; ZIP: 83415-3810

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

; COMPUTER: Toshiba Satellite Pro T2150CDS

; OPERATING SYSTEM: Windows95

; SOFTWARE: Word Perfect 7.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/064,693A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: W. Gary Goodson

; REGISTRATION NUMBER: 22,387

; REFERENCE/DOCKET NUMBER: LIT-PI-296

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (208)526-9469

; TELEFAX: (208)526-8339

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1348 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-09-064-693A-22

Query Match 61.9%; Score 19.2; DB 4; Length 1348;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 CTTTACATCCCGCTTCAATG 31
|||
Db 830 CTTTACAGTCCGCTTGAATG 807

RESULT 3

US-09-064-693A-26/C

; Sequence 26, Application US/09064693A

; Patent No. 6210937

; GENERAL INFORMATION:

; APPLICANT: Ward, Thomas E.

; TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY

TITLE OF INVENTION: ENGINEERED BACTERIA FOR PRODUCTION
TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: W. Gary Goodson

; ADDRESSEE: INEL--Lockheed Martin Idaho

; ADDRESSEE: Technologies Co.

; STREET: P.O. Box 1625

; CITY: Idaho Falls

; STATE: Idaho

; COUNTRY: USA

; ZIP: 83415-3810

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage

; COMPUTER: Toshiba Satellite Pro T2150CDS

; OPERATING SYSTEM: Windows95

; SOFTWARE: Word Perfect 7.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/064,693A

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: W. Gary Goodson

; REGISTRATION NUMBER: 22,387

; REFERENCE/DOCKET NUMBER: LIT-PI-296

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (208)526-9469

; TELEFAX: (208)526-8339

; INFORMATION FOR SEQ ID NO: 26:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4530 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

US-09-064-693A-26

Query Match 61.9%; Score 19.2; DB 4; Length 4530;
Best Local Similarity 87.5%; Pred. No. 25;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 CTTTACATCCCGCTTCAATG 31
|||
Db 2712 CTTTACAGTCCCGCTTGAATG 2689

RESULT 4

US-08-961-527-195

; Sequence 195, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 195:
SEQUENCE CHARACTERISTICS:
LENGTH: 6693 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-195

Query Match 60.6%; Score 18.8; DB 4; Length 6693;
Best Local Similarity 76.7%; Pred. No. 39;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTACATTCGCCCTTCAAT 30
Db 492 ATTCCACCTTTGATGACCGCTTGAGTT 521

RESULT 5

US-08-619-812-5/C
Sequence 5, Application US/08619812
Patent No. 610066

GENERAL INFORMATION:

APPLICANT: POTTER, ANDREW A.

APPLICANT: THEISEN, MICHAEL

APPLICANT: HARLAND, RICHARD J.

APPLICANT: RIOUX, CLEMENT R.

TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: REED & ROBINS

STREET: 635 BRYANT STREET

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/619,812

FILING DATE: 15-MAR-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/038,719

FILING DATE: 29-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: ROBINS, ROBERTA L.

REGISTRATION NUMBER: 33,208

REFERENCE/DOCKET NUMBER: 9000-0019.20

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 617-8999

TELEFAX: (415) 327-3231

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2465 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 108..1850

US-08-619-812-5

Query Match 59.4%; Score 18.4; DB 3; Length 2465;
Best Local Similarity 78.6%; Pred. No. 48;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTACATTCGCCCTTTCA 28
Db 1220 ATTCCACATTTTACCTGCGCACCTTCA 1193

RESULT 6

US-08-619-554-3/C
Sequence 3, Application US/08619554
Patent No. 5821353

GENERAL INFORMATION:

APPLICANT: DOUGLAS, Cameron M.

APPLICANT: CHREBET, Gary L.

APPLICANT: CLEWAS, Joseph

APPLICANT: EL-SHERBEINI, Mohammed

APPLICANT: FOOR, Forrest

APPLICANT: KAHN, Jennifer,

APPLICANT: KELLY, Rosemarie, - PARENT, S.A.

APPLICANT: MARRINAN, Jean, - RAMADAN, N.M.

APPLICANT: MORIN, Nancy, - REGISTE, E.A

APPLICANT: ONISHI, Janet, - SHEI, Gan-Ju

TITLE OF INVENTION: DNA ENCODING 1,3 BETA-D GLUCAN

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: JOSEPH A. COPPOLA - MERCK & CO., INC.

STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000

CITY: RAHWAY

STATE: NJ

COUNTRY: USA

ZIP: 07065

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/619,554

FILING DATE: 01-AUG-1996

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: COPPOLA, JOSEPH A

REGISTRATION NUMBER: 38,413

REFERENCE/DOCKET NUMBER: 19104PI

TELECOMMUNICATION INFORMATION:

TELEPHONE: 732-594-6734

TELEFAX: 732-594-4720

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 7070 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-619-554-3

Query Match 58.7%; Score 18.2; DB 1; Length 7070;
Best Local Similarity 74.2%; Pred. No. 70;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTACATTCGCCCTTCAATG 31
Db 46 ATTCCACCTTTGACATGCGCATCTTCAATG 16

RESULT 7
US-09-071-035-147/c
; Sequence 147, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071.035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 147:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 298 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-071-035-147
Query Match 58.1%; Score 18; DB 4; Length 298;
Best Local Similarity 80.8%; Pred. No. 48;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy 1 ATTCACCTTTACATTCGCCCTTT 26
Db 249 ATTACACCTTTAATTACCGCTTT 224
RESULT 8
US-09-071-035-145/c
; Sequence 145, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071.035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 145:
SEQUENCE CHARACTERISTICS:
LENGTH: 441 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-145
Query Match 58.1%; Score 18; DB 4; Length 441;
Best Local Similarity 80.8%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy 1 ATTCACCTTTACATTCGCCCTTT 26
Db 392 ATTACACCTTTAATTACCGCTTT 367
RESULT 9
US-08-961-083-137/c
; Sequence 137, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 137:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-083-137
Query Match 58.1%; Score 18; DB 3; Length 502;

Best Local Similarity 80.8%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 CCACCTTTACATCCGCGCTTTCAT 29
Db 420 CCACCTTGAGATGTCGCGCTTTCAT 395

RESULT 10

US-08-961-527-48/C
; Sequence 48, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-48

Query Match 58.1%; Score 18; DB 4; Length 25002;
Best Local Similarity 80.8%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 CCACCTTTACATCCGCGCTTTCAT 29
Db 10612 CCACCTTGAGATGTCGCGCTTTCAT 10587

RESULT 11

US-09-328-111-101/C
; Sequence 101, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Asie, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 685
; TYPE: DNA
; ORGANISM: Homo sapiens

; NAME/KEY: misc_feature
; LOCATION: (1)...(685)
; OTHER INFORMATION: n = A,T,C or G
; US-09-328-111-101

Query Match 57.4%; Score 17.8; DB 4; Length 685;
Best Local Similarity 82.6%; Pred. No. 67;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 CCTTTACATCCGCGCTTTCAT 29
Db 527 CCNTGNCATTCGCCCTTTCAT 505

RESULT 12

US-09-397-885-2/C
; Sequence 2, Application US/09397885
; Patent No. 6323007
; GENERAL INFORMATION:
; APPLICANT: Moller, Soren
; APPLICANT: Johansen, Charlotte
; APPLICANT: Schaefer, Thomas
; APPLICANT: Ostergaard, Peter Rahbek
; APPLICANT: Hoeck, Lisbeth Hedegaard
; TITLE OF INVENTION: A 2,6, -b-D-Fructan Hydrolase Enzyme And
; FILE REFERENCE: 5540.200-US
; CURRENT APPLICATION NUMBER: US/09/397,885
; EARLIER FILING DATE: 1999-09-17
; EARLIER APPLICATION NUMBER: PA 1998 01173
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: PA 1998 01623
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 60/101,615
; EARLIER FILING DATE: 1998-09-24
; EARLIER APPLICATION NUMBER: 60/111,675
; EARLIER FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2769
; TYPE: DNA
; ORGANISM: Paenibacillus amylolyticus
; US-09-397-885-2

Query Match 56.8%; Score 17.6; DB 4; Length 2769;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 CACCTTTACATCCGCGCTTTCAT 28
Db 2584 CACATTCACATTCGCCCTTTTA 2561

RESULT 13

US-09-058-489-29

Sequence 29, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
FILE REFERENCE: WH197-089A
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 29
LENGTH: 1202
TYPE: DNA
ORGANISM: Human
US-09-058-489-29

Query Match 56.1%; Score 17.4; DB 3; Length 1202;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATTCACCTTTACATTCCTCCGCTTTC 27
DB 648 ACTCAACATTTACATTCATCTTTC 674

RESULT 14
US-08-938-669A-1/c
Sequence 1, Application US/08938669A
Patent No. 6171788
GENERAL INFORMATION:
APPLICANT: Nguyen, Thai D.
APPLICANT: Polansky, Jon R.
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS,
PROGNOSIS AND TREATMENT OF GLAUCOMA AND
TITLE OF INVENTION: RELATED DISEASES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004-2402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,669A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/791,154
FILING DATE: 28-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mendelson, Elliot
REGISTRATION NUMBER: P-42,878
REFERENCE/DOCKET NUMBER: 07425-0034
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 383-6857
TELEFAX: 202 383-6610
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5300 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
US-08-938-669A-1

Query Match 56.1%; Score 17.4; DB 4; Length 5300;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTCACCTTTACATTCCTCCGCTTCA 28
DB 2573 TTGCGCTTTACCTTCCTGCTTCA 2547

RESULT 15
US-08-938-669A-2/c
Sequence 2, Application US/08938669A
Patent No. 6171788
GENERAL INFORMATION:
APPLICANT: Nguyen, Thai D.
APPLICANT: Polansky, Jon R.
TITLE OF INVENTION: METHODS FOR THE DIAGNOSIS,
PROGNOSIS AND TREATMENT OF GLAUCOMA AND
TITLE OF INVENTION: RELATED DISEASES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004-2402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,669A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/791,154
FILING DATE: 28-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Mendelson, Elliot
REGISTRATION NUMBER: P-42,878
REFERENCE/DOCKET NUMBER: 07425-0034
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 383-6857
TELEFAX: 202 383-6610
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5304 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-938-669A-2

Query Match 56.1%; Score 17.4; DB 4; Length 5304;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTCACCTTTACATTCCTCCGCTTCA 28
DB 2573 TTGCGCTTTACCTTCCTGCTTCA 2547

Search completed: May 11, 2003, 03:08:05
Job time : 16.4311 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:08:16 ; Search time 30.6892 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-17

Perfect score: 31

Sequence: 1 attccacattacattccgccttcattg 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCRUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	31	9 US-09-963-803-17	Sequence 17, Appl
2	20.6	66.5	1691139	9 US-10-067-514-1	Sequence 1, Appl
3	19.8	63.9	387	10 US-09-960-352-14643	Sequence 14643, A
4	19.8	63.9	498	10 US-09-964-824A-23	Sequence 23, Appl
5	19.4	62.6	291	10 US-09-974-300-3911	Sequence 3911, Ap
6	19.4	62.6	595	10 US-09-864-761-15272	Sequence 15272, A
7	19.4	62.6	1671	10 US-09-815-242-6926	Sequence 6926, Ap
8	19.4	62.6	2000	9 US-09-938-842A-4351	Sequence 4351, Ap
9	19.4	62.6	501	10 US-09-783-590-526	Sequence 526, Ap
10	18.8	60.6	493	9 US-09-796-682-8653	Sequence 8653, Ap
11	18.8	60.6	493	9 US-10-040-862-8653	Sequence 8653, Ap
12	18.8	60.6	1038	10 US-09-822-849A-431	Sequence 431, Ap
13	18.8	60.6	32192	9 US-09-764-891-9791	Sequence 9791, Ap
14	18.8	60.6	32192	10 US-09-764-877-3657	Sequence 3657, Ap
15	18.6	60.0	32768	10 US-09-070-927A-399	Sequence 399, Ap
16	18.4	59.4	417	10 US-09-974-300-611	Sequence 611, Ap
17	18.4	59.4	468	9 US-10-079-854-96	Sequence 96, Appl
18	18.4	59.4	468	10 US-09-764-878-96	Sequence 96, Appl
19	18.4	59.4	580	10 US-09-560-863-525	Sequence 525, Appl

20	18.4	59.4	1039	7 US-08-781-986A-273	Sequence 273, App
21	18.4	59.4	1057	10 US-09-974-300-249	Sequence 249, App
22	18.2	58.7	1088	10 US-09-974-300-7471	Sequence 7471, Ap
23	18.2	58.7	293	10 US-09-294-093B-3209	Sequence 3209, Ap
24	18.2	58.7	5230	10 US-09-838-718A-5	Sequence 5, Appl
25	18.2	58.7	5770	10 US-09-838-718A-7	Sequence 7, Appl
26	18.2	58.7	5870	10 US-09-838-718A-8	Sequence 8, Appl
27	18.2	58.7	5906	10 US-09-838-718A-6	Sequence 6, Appl
28	18.2	58.7	9412	9 US-10-032-393-51	Sequence 51, Appl
29	18.2	58.7	9412	9 US-10-032-393-52	Sequence 52, Appl
30	18.2	58.7	9417	9 US-10-032-393-48	Sequence 48, Appl
31	18.2	58.7	9425	9 US-10-032-393-46	Sequence 46, Appl
32	18.2	58.7	9437	9 US-10-032-393-17	Sequence 17, Appl
33	18.2	58.7	9484	9 US-10-032-393-49	Sequence 49, Appl
34	18.2	58.7	9566	9 US-10-032-393-50	Sequence 50, Appl
35	18.2	58.7	10929	9 US-10-032-393-7	Sequence 7, Appl
36	18.2	58.7	12733	9 US-10-032-393-47	Sequence 47, Appl
37	18.2	58.7	12739	9 US-10-032-393-8	Sequence 8, Appl
38	18.2	58.7	274	10 US-09-878-574-13572	Sequence 13572, A
39	18.2	58.1	502	10 US-09-765-272-137	Sequence 137, App
40	18.2	58.1	612	10 US-09-070-927A-832	Sequence 832, App
41	18.2	58.1	724	10 US-09-070-927A-931	Sequence 931, App
42	17.8	57.4	272	10 US-09-878-574-8758	Sequence 8758, Ap
43	17.8	57.4	362	9 US-09-796-692-6939	Sequence 6939, Ap
44	17.8	57.4	362	9 US-09-796-692-9539	Sequence 9539, Ap
45	17.8	57.4	362	9 US-10-040-862-6999	Sequence 6999, Ap

ALIGNMENTS

RESULT 1
US-09-963-803-17
Sequence 17, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 17
LENGTH: 31
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: block G3
US-09-963-803-17

Query Match 100.0%; Score 31; DB 9; Length 31;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 ATTCACCTTTACATTCGCCCTTTCAATG 31
1 ATTCACCTTTACATTCGCCCTTTCAATG 31

RESULT 2
US-10-067-514-1/c
Sequence 1, Application US/10067514
Publication No. US20030054531A1
GENERAL INFORMATION:
APPLICANT: Gileadscotlr, Solveig
APPLICANT: Jonsdottr, Sif

APPLICANT: Revnisdottir, Sigridur Th.
TITLE OF INVENTION: HUMAN STROKE GENE
FILE REFERENCE: 2345.2010-003
CURRENT APPLICATION NUMBER: US/10/067,514
CURRENT FILING DATE: 2002-02-04
PRIOR APPLICATION NUMBER: US 09/811/352
PRIOR FILING DATE: 2001-03-19
NUMBER OF SEQ ID NOS: 84
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1691139
TYPE: DNA
ORGANISM: Human
US-10-067-514-1

Query Match 66.5%; Score 20.6; DB 9; Length 1691139;
Best Local Similarity 85.2%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 4 CCACCTTTACATTCGCGCTTCATTG 30
Db 929591 CCACCTTTACATTCGCGCTTCATTG 929595

RESULT 3
US-09-960-352-14643
Sequence 14643, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Teo, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 14643
LENGTH: 387
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 62-LIB34-082-Q1-E1-H10
US-09-960-352-14643

Query Match 63.9%; Score 19.8; DB 10; Length 387;
Best Local Similarity 77.4%; Pred. No. 66;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTACATTCGCGCTTCATTG 31
Db 39 ATTCCACCTTTACATTCGCGCTTCATTG 69

RESULT 4
US-09-964-824A-23/C
Sequence 23, Application US/09964824A
Patent No. US20020102531A1
GENERAL INFORMATION:
APPLICANT: Horigan, Stephen
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
FILE REFERENCE: 689290-71
CURRENT APPLICATION NUMBER: US/09/964,824A
CURRENT FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: US/60/236,033
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,032
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US/60/236,028
PRIOR FILING DATE: 2000-09-28
NUMBER OF SEQ ID NOS: 583

SOFTWARE: PatentIn version 3.0
SEQ ID NO 23
LENGTH: 498
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(498)
OTHER INFORMATION: n=a,t,g or c
US-09-964-824A-23

Query Match 63.9%; Score 19.8; DB 10; Length 498;
Best Local Similarity 77.4%; Pred. No. 69;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATTCCACCTTTACATTCGCGCTTCATTG 31
Db 158 ATTCCACCTTTACATTCGCGCTTCATTG 128

RESULT 5
US-09-974-300-3911
Sequence 3911, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods for Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3911
LENGTH: 291
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-3911

Query Match 62.6%; Score 19.4; DB 10; Length 291;
Best Local Similarity 79.3%; Pred. No. 92;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 TTCCACCTTTACATTCGCGCTTCATTG 30
Db 256 TTCCACCTTTACATTCGCGCTTCATTG 284

RESULT 6
US-09-864-761-15272/C
Sequence 15272, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03

```
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15272
; LENGTH: 595
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018469.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER. SIGNAL = 0.66
; OTHER INFORMATION: EXPRESSED IN PLACENTA. SIGNAL = 0.51
US-09-864-761-15272

Query Match          62.6%; Score 19.4; DB 10; Length 595;
Best Local Similarity 79.3%; Pred. No. 1e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCCTTTCATT 30
Db 64 TTCCACCTTTTACATTCCTGCTTTT 36

RESULT 7
US-09-815-242-6926/c
; Sequence 6926, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zysek, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6926
; LENGTH: 1671
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1671)
US-09-815-242-6926

Query Match          62.6%; Score 19.4; DB 10; Length 1671;
Best Local Similarity 79.3%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCCTTTCATT 30
Db 421 TTTCACCTTTTACATTCACACTTTCCTT 393

RESULT 8
US-09-938-842A-4351/c
; Sequence 4351, Application US/0938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krepes, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,966
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4351
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4351

Query Match          62.6%; Score 19.4; DB 9; Length 2000;
Best Local Similarity 79.3%; Pred. No. 1.2e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCCTTTCATT 30
Db 1704 TTCCACCTTTTACATTCCTGCTTTT 1676

RESULT 9
US-09-783-590-526/c
; Sequence 526, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
```

```

APPLICANT: Li, Haodong
APPLICANT: Rosen, Craig A.
APPLICANT: Ruben, Steven M.
TITLE OR INVENTION: Human Genes, Sequences, and Expression Products 16.2.
FILE REFERENCE: P0-16.2c1
CURRENT APPLICATION NUMBER: US/09/783,590
PRIORITY FILING DATE: 2000-02-15
PRIORITY APPLICATION NUMBER: 08/420,856
PRIORITY FILING DATE: 1995-04-12
PRIORITY APPLICATION NUMBER: 08/446,731
PRIORITY FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12485
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 526
LENGTH: 501
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (74)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (211)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (243)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (277)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (290)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (322)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (327)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (331)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (338)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (345)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (350)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (351)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (366)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (374)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (375)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (402)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (405)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (413)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature

```

```

1 LOCATION: (416)
2 OTHER INFORMATION: n equals a,t,g, or c
3 NAME/KEY: misc feature
4 LOCATION: (420)
5 OTHER INFORMATION: n equals a,t,g, or c
6 NAME/KEY: misc feature
7 LOCATION: (421)
8 OTHER INFORMATION: n equals a,t,g, or c
9 NAME/KEY: misc feature
10 LOCATION: (423)
11 OTHER INFORMATION: n equals a,t,g, or c
12 NAME/KEY: misc feature
13 LOCATION: (430)
14 OTHER INFORMATION: n equals a,t,g, or c
15 NAME/KEY: misc feature
16 LOCATION: (433)
17 OTHER INFORMATION: n equals a,t,g, or c
18 NAME/KEY: misc feature
19 LOCATION: (438)
20 OTHER INFORMATION: n equals a,t,g, or c
21 NAME/KEY: misc feature
22 LOCATION: (444)
23 OTHER INFORMATION: n equals a,t,g, or c
24 NAME/KEY: misc feature
25 LOCATION: (452)
26 OTHER INFORMATION: n equals a,t,g, or c
27 NAME/KEY: misc feature
28 LOCATION: (455)
29 OTHER INFORMATION: n equals a,t,g, or c
30 NAME/KEY: misc feature
31 LOCATION: (460)
32 OTHER INFORMATION: n equals a,t,g, or c
33 NAME/KEY: misc feature
34 LOCATION: (463)
35 OTHER INFORMATION: n equals a,t,g, or c
36 NAME/KEY: misc feature
37 LOCATION: (476)
38 OTHER INFORMATION: n equals a,t,g, or c
39 NAME/KEY: misc feature
40 LOCATION: (480)
41 OTHER INFORMATION: n equals a,t,g, or c
42 NAME/KEY: misc feature
43 LOCATION: (484)
44 OTHER INFORMATION: n equals a,t,g, or c
45 US-09-783-590-526
46
47 Query Match 61.3%, Score 19; DB 10; Length 501;
48 Best Local Similarity 75.9%; Pred. No. 1.4e+02;
49 Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
50
51 QY 2 TTCACCTTTTACATTCGCCCTTTCATT 30
52 ||||| ||||| ||||| ||||| |||||
53
54 Db 236 TTCACGNTTTAACTGCACGCTTCGATT 268
55
56 RESULT 10
57 US-09-796-692-8653/C
58 ; Sequence 8653, Application US/09796692
59 ; Publication No. US20020198362v1
60 ; GENERAL INFORMATION:
61 ; APPLICANT: Gaiger, Alexander
62 ; APPLICANT: Algate, Paul A.
63 ; APPLICANT: Mannion, Jane
64 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAP
65 ; FILE REFERENCE: 2077.001200
66 ; CURRENT APPLICATION NUMBER: US/09/796,692
67 ; CURRENT FILING DATE: 2001-03-01
68 ; PRIOR APPLICATION NUMBER: 60/186,126
69 ; PRIOR FILING DATE: 2000-03-01
70 ; PRIOR APPLICATION NUMBER: 60/190,479
71 ; PRIOR FILING DATE: 2000-03-17
72 ; PRIOR APPLICATION NUMBER: 60/200,545

```

PRIOR FILING DATE: 2000-04-27
 PRIOR APPLICATION NUMBER: 60/200,303
 PRIOR FILING DATE: 2000-04-28
 PRIOR APPLICATION NUMBER: 60/200,779
 PRIOR FILING DATE: 2000-04-28
 PRIOR APPLICATION NUMBER: 60/200,999
 PRIOR FILING DATE: 2000-05-01
 PRIOR APPLICATION NUMBER: 60/202,084
 PRIOR FILING DATE: 2000-05-04
 PRIOR APPLICATION NUMBER: 60/206,201
 PRIOR FILING DATE: 2000-05-22
 PRIOR APPLICATION NUMBER: 60/218,950
 PRIOR FILING DATE: 2000-07-14
 PRIOR APPLICATION NUMBER: 60/222,903
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: 60/223,416
 PRIOR FILING DATE: 2000-08-04
 PRIOR APPLICATION NUMBER: 60/223,378
 PRIOR FILING DATE: 2000-08-07
 NUMBER OF SEQ ID NOS: 9597
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 8653
 LENGTH: 493

TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (117)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (158)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (165)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (225)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (234)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (252)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (279)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (308)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (332)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (386)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (389)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (465)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (486)
 OTHER INFORMATION: n=A,T,C or G
 US-09-796-692-8653

Query Match 60.6%; Score 18.8; DB 9; Length 493;
 Best local Similarity 74.2%; Pred. No. 1.7e+02;
 Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1 ATTCCACCTTTTACATTCGCCCTTTTCATG 31

Db 226 ANACATTTTAAAGTCCACCTTTTCATG 196

RESULT 11
 US-10-040-862-8653/C
 Sequence 8653, Application US/10040862
 Publication No. US20030078396A1
 GENERAL INFORMATION:
 APPLICANT: Gaiger, Alexander
 APPLICANT: Algate, Paul A.
 APPLICANT: Mannion, Jane
 APPLICANT: Retter, Marc
 APPLICANT: Corixa Corporation
 TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy of Hematological Malignancies
 FILE REFERENCE: 014058-013520US
 CURRENT APPLICATION NUMBER: US/10/040,862
 CURRENT FILING DATE: 2001-11-06
 PRIOR APPLICATION NUMBER: US 60/186,126
 PRIOR FILING DATE: 2000-03-01
 PRIOR APPLICATION NUMBER: US 60/190,479
 PRIOR FILING DATE: 2000-03-17
 PRIOR APPLICATION NUMBER: US 60/200,545
 PRIOR FILING DATE: 2000-04-27
 PRIOR APPLICATION NUMBER: US 60/200,303
 PRIOR FILING DATE: 2000-04-28
 PRIOR APPLICATION NUMBER: US 60/200,779
 PRIOR FILING DATE: 2000-04-28
 PRIOR APPLICATION NUMBER: US 60/200,999
 PRIOR FILING DATE: 2000-05-01
 PRIOR APPLICATION NUMBER: US 60/202,084
 PRIOR FILING DATE: 2000-05-04
 PRIOR APPLICATION NUMBER: US 60/206,201
 PRIOR FILING DATE: 2000-05-22
 PRIOR APPLICATION NUMBER: US 60/218,950
 PRIOR FILING DATE: 2000-07-14
 PRIOR APPLICATION NUMBER: US 60/222,903
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: US 60/223,416
 PRIOR FILING DATE: 2000-08-04
 PRIOR APPLICATION NUMBER: US 60/223,378
 PRIOR FILING DATE: 2000-08-07
 PRIOR APPLICATION NUMBER: US 09/796,692
 PRIOR FILING DATE: 2001-03-01
 NUMBER OF SEQ ID NOS: 10467
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 8653
 LENGTH: 493

TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (117)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (158)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (225)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (234)
 OTHER INFORMATION: n=A,T,C or G
 NAME/KEY: unsure
 LOCATION: (252)

```
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (279)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (308)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (332)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (386)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (389)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (465)
OTHER INFORMATION: n=A,T,C or G
FEATURE:
NAME/KEY: unsure
LOCATION: (486)
OTHER INFORMATION: n=A,T,C or G
US-10-040-862-8653
```

Query Match 60.6%; Score 18.8; DB 9; Length 493;

Best Local Similarity 74.2%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
OY 1 ATTCACCTTTTACATTCGCGCTTTCATTG 31
Db 226 AMACCATTTTAAAGTCCACCTTTCATTG 196
```

```
RESULT 12
US-09-822-849A-431/C
Sequence 431, Application US/09822849A
Patent No. US20020045170A1
GENERAL INFORMATION:
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fichtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulikota, Kamalakari
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
FILE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
TITLE REFERENCE: GIN 6403
CURRENT APPLICATION NUMBER: US/09/822, 849A
CURRENT FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/195,582
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 598
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 431
LENGTH: 1038
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-849A-431
```

Query Match 60.6%; Score 18.8; DB 10; Length 1038;
Best Local Similarity 76.7%; Pred. No. 1.9e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 ATTCACCTTTTACATTCGCGCTTTCATT 30

```
Db 519 AATCCCATTTTCCTTCCACCTTCATT 490
```

```
RESULT 13
US-09-764-891-9791/C
Sequence 9791, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9791
LENGTH: 32192
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-9791
```

Query Match 60.6%; Score 18.8; DB 9; Length 32192;
Best Local Similarity 76.7%; Pred. No. 3.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
OY 1 ATTCACCTTTTACATTCGCGCTTTCATT 30
Db 955 ATTCTAATTTCTACTTCCAGCTTACATT 926
```

```
RESULT 14
US-09-764-877-3657
Sequence 3657, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3657
LENGTH: 32192
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-877-3657
```

Query Match 60.6%; Score 18.8; DB 10; Length 32192;
Best Local Similarity 76.7%; Pred. No. 3.3e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
OY 1 ATTCACCTTTTACATTCGCGCTTTCATT 30
Db 31238 ATTCTAATTTCTACTTCCAGCTTACATT 31267
```

```
RESULT 15
US-09-070-927A-399/C
Sequence 399, Application US/09070927A
Patent No. US20020120116A1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
APPLICANT: Patrick J. Dillon
APPLICANT: Steven Barash
TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
```

```

: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/070.927A
: FILING DATE: 04-May-2000
: CLASSIFICATION: <Unknown>
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/046,655
: FILING DATE: 1997-05-16
: APPLICATION NUMBER: 60/044,031
: FILING DATE: 1997-05-06
: APPLICATION NUMBER: 60/066,009
: FILING DATE: 1997-11-14
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Kenley K. Hoover
: REGISTRATION NUMBER: 40,302
: REFERENCE/DOCKET NUMBER: PB369
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 399:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 32768 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: US-09-070-927A-399

```

```

Query Match      60.0%; Score 18.6; DB 10; Length 32768;
Best Local Similarity 84.0%; Pred. No. 4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATTGCACCTTTTACATTCGCCGCTT 25
||||| ||||| ||||| |||||
Db 30220 ATTGCACCTTTTACATTCGCCGCTT 30196

```

Search completed: May 11, 2003, 06:31:22
 Job time : 231.689 secs

Tel: 515-294-0975
 Fax: 515-294-2299
 Email: schnablab@iastate.edu
 Individual basical and confidence value were assigned using the
 phred software,
 (<http://depts.washington.edu/ventures/collabtr/direct/index.htm#btr>). Overall sequence quality assessment and vector trimming were
 conducted using the Lucy software (<http://www.tigr.org/sofclab/>).
 Lucy parameters were set to ensure an overall trimmed quality of
 97.5% or better without any vector fragments in the chosen
 high-quality region of each sequence. Low-quality bases between the
 poly-T and the high-quality region were replaced with N's to serve
 as spacers.
 PCR Primers
 FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
 BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)
 Seq primer: universal (GTA AAA CGA CGG CCA GT).
 Location/Qualifiers
 1. 599
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone="MEST407-A04"
 /clone_11b="ISUM5-RN"
 /tissue_type="mixed"
 /lab_host="DH10B"
 /note="Vector: pT73PAC; Site 1: EcoRI; Site 2: NotI;
 Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),
 Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels
 (3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65 DAG
), Tassel (3-39 cm, 53 and 56 DAG), Immature ear (0.2-3.0
 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated
 first ear, ear shank, etiolated seedlings, callus,
 Cycloheximide-treated callus, anaerobic treated seedlings
 , NAA (a-Naphthalene acetic acid)-treated seedlings,
 Kinetin-treated seedlings, ACPc
 (1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
 Brassinolide-treated seedlings, ABA (Absciscic acid
)-treated seedlings, GA (Gibberellic acid)-treated
 seedlings, JA (Jasmonic acid)-treated seedlings, ds-cDNA
 molecules were generated as follows. First-strand cDNA was
 prepared from oligo-dT selected mRNA by priming with a
 NotI oligo-dT primer (5',
 ACTGGAAGATTGCGCGCCGAGATTTTCTTTT). The
 resulting DNA:RNA hybrid was treated with RNase H and used
 as a template for DNA Pol-I-catalyzed second strand
 synthesis. After the addition of EcoRI adaptor, the
 ds-cDNAs were digested with NotI and size-selected. The
 resulting molecules were directionally cloned into the
 EcoRI and NotI sites of the pT73PAC vector. The library
 then went through one round of normalization to COT value
 of 5 based on the methods of Marcelo Bento Soares (Genome
 Research 6: 791-806, 1996)."

BASE COUNT 159 a 147 c 115 g 178 t

Query Match 71.0%; Score 22; DB 13; Length 599;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 2 TTCACCTTTACATTCGCGCTTCATTG 31
 518 TTACACCTTAACATCCCGCCTTAATTG 547

RESULT 4 BM317907 599 bp mRNA linear EST 03-JAN-2002
 LOCUS BM317907/1
 DEFINITION P11.13_A02.g9_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION BM317907
 VERSION BM317907.1 GI:18052249
 KEYWORDS EST.

SOURCE
 ORGANISM
 Sorghum
 Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 599)
 Cordonnier-Pratt M.-M., Gingle A., Dean R., Sudman M. and Pratt
 L.H.
 An EST database from Sorghum: pathogen-induced plants
 Unpublished (2000)
 Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude polyA, vector and regions
 below phred quality 16. The threshold for highest quality sequence
 is 20. Three-prime sequences, which are obtained with POLYmix or
 T7 sequencing primer, are presented as the reverse complement.
 Seq primer: T7
 High quality sequence start: 60
 High quality sequence stop: 599
 POLYA=Yes.
 Location/Qualifiers
 1. 599
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_11b="Pathogen induced 1 (P11)"
 /note="Organ: Anthracnose-infected leaves from
 two-week-old sorghum plants 48 hr after inoculation;
 Vector: pBluescript II from Lambda Zap II; Site 1: XhoI;
 Site 2: EcoRI; Two-week-old sorghum plants (BRX 623
 cultivar) were infected with pathogen (isolate FRW421 of
 Colletotrichum graminicola, which is a sorghum isolate).
 RNA was prepared from infected leaves harvested from 45
 seedlings 48 hours after inoculation. Note: Young
 seedlings (2 weeks old) exhibit juvenile resistant
 reaction, which is an incompatible interaction. As they
 grow older (4 weeks or older), plants resume susceptibility
 to anthracnose disease. The library was made from poly-A
 RNA in the cloning vector lambda Zap II. Clones to be
 sequenced were prepared by mass excision. WARNING: While
 most or all ESTs are expected to derive from the host
 plant, no effort was made to eliminate ESTs deriving from
 the pathogen."

BASE COUNT 170 a 117 c 142 g 170 t

Query Match 71.0%; Score 22; DB 13; Length 599;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 2 TTCACCTTTACATTCGCGCTTCATTG 31
 44 TTACACCTTAACATCCCGCCTTAATTG 15

RESULT 5 BM952936 614 bp mRNA linear EST 14-MAR-2002
 LOCUS BM952936/1
 DEFINITION 952060C11.X1.952 - BMS tissue from Walbot lab (reduced RNA) Zea
 mays cDNA, mRNA sequence.
 ACCESSION BM952936
 VERSION BM952936.1 GI:19436526
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 614)

AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 952060 row: C column: 11.

FEATURES
Source
 1. .614
 /organism="Zea mays"
 /cultivar="BMS (Black Mexican Sweet)"
 /db_xref="taxon:4577"
 /clone_lib="952 - BMS tissue from Walbot Lab (reduced RNA)"
 /issue_type="suspension culture"
 /dev_stage="mixed logarithmic and stationary growth phases"
 /lab_host="DH10B"
 /note="Vector: pUC19, Site_1: EcoRI, Site_2: EcoRI; The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive clones."

BASE COUNT 175 a 144 c 150 g 145 t

ORIGIN

Query Match 71.0%; Score 22; DB 14; Length 614;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 TTCACCTTTACATTCCTTCATTG 31
 |||||
 Db 567 TTACACCTTACATCCCGCTTATTG 538

RESULT 6
LOCUS BM330506/c 631 bp mRNA linear EST 04-JAN-2002
DEFINITION PIC1 52_H06.g1_A002 Pathogen-infected compatible 1 (PIC1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BM330506
VERSION BM330506.1 GI:18069643
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 631)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Fang, G.C., Dean, R., Wing, R., Sudman, M. and Pratt, L.H.
TITLE An EST database from Sorghum: plants infected with a compatible pathogen
JOURNAL Unpublished (2002)
COMMENT Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmp@pratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector, and regions below Phred quality 16. The threshold for highest quality sequence

is 20. Three-prime sequences, which are obtained with PolyTmX or T7 sequencing primer, are presented as the reverse complement.
 Seg primer: T7
 High quality sequence start: 100
 High quality sequence stop: 631
 PolyA=yes.

FEATURES
Source
 1. .631
 /organism="Sorghum bicolor"
 /cultivar="Brx623"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen-infected compatible 1 (PIC1)"
 /issue_type="Leaves"
 /dev_stage="4-week-old seedlings infected with Colletotrichum graminicola"
 /note="Vector: pBluescript II SK(-) from Lambda Zap II; Site_1: XhoI, Site_2: EcoRI; Four-week-old sorghum seedlings were sprayed with spore suspension prepared from 3-week-old PM421, a sorghum isolate of the anthracnose pathogen Colletotrichum graminicola. Inoculated plants were kept in a 25 C dark growth chamber with 100% relative humidity for 24 hr, followed by 12/12 hr of light/dark cycle at 25 C with 90% relative humidity for another 24 hr. All leaves were harvested and quick frozen with liquid nitrogen and stored in a -80 C freezer. The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 181 a 120 c 154 g 176 t

ORIGIN

Query Match 71.0%; Score 22; DB 13; Length 631;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 TTCACCTTTACATTCCTTCATTG 31
 |||||
 Db 94 TTACACCTTACATCCCGCTTATTG 65

RESULT 7
LOCUS BQ283489/c 639 bp mRNA linear EST 13-MAY-2002
DEFINITION WHE3091.G03 M05ZS Wheat cold-stressed seedling subtracted cDNA library Triticum aestivum cDNA clone WHE3091_G03_M05, mRNA sequence.
ACCESSION BQ283489
VERSION BQ283489.1 GI:20553240
KEYWORDS EST.
SOURCE bread wheat
ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticaceae; Triticum.
REFERENCE 1 (bases 1 to 639)
AUTHORS Anderson, O.D., Cho, S., Choi, D.W., Close, T.J., Crossman, C., Penton, R.D., Izzo, G.R., Nguyen, H.T., Phan, J., Rauech, C.J., Wilson, C., Woo, J. and Zhang, D.
TITLE The structure and function of the expressed portion of the wheat genomes - Cold-stressed seedling subtracted cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.ars.gov
 Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
 Seg primer: SK primer.

FEATURES

source

Location/Qualifiers
1. .639
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="MH83091_G03_M05"
/clone_1lb="Wheat cold-stressed seedling subtracted cDNA library"

/tissue type="Seedling"
/dev stage="Five-day old seedling"
/lab_host="E. coli SOLr"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water, mycstatin and cefotaxime in covered crystallization dishes. Five-day old seedlings were transferred to 5 C cold room and kept for 48 hr. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give Bluescript phagemids in the TJ Close lab (Choi, Close, Penion) at the University of California, Riverside. The cDNA clones were in vivo excised to give Bluescript phagemids before subtraction was carried out. The mass excision of phagemid library and subtraction were done in HT Nguyen lab by D. Zhang at Texas Tech University. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)"

BASE COUNT 192 a 132 c 157 g 157 t 1 others
ORIGIN

Query Match 71.0%; Score 22; DB 14; Length 639;
Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATCCCGCCTTCAATG 31

Db 301 TTACACCTTAACATCCCGCCTTCAATG 272

RESULT 8
BM078660
LOCUS

DEFINITION BM078660 644 bp mRNA linear EST 14-NOV-2001
MEST122-F07.T3 ISUM4-TN Zea mays cDNA clone MEST122-F07 3', mRNA

ACCESSION BM078660
VERSION BM078660.1 GI:16925592

KEYWORDS
SOURCE
ORGANISM

Zea mays.
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.

REFERENCE

AUTHORS Qiu, F., Cui, F., Guo, L., Ashlock, D.A., Wen, T.J. and Schnable, P.S.
TITLE Expressed Sequence Tags from B73 Maize Seedlings and Silks
JOURNAL Unpublished (2001)

COMMENT Contact: Patrick S. Schnable
Schnable Laboratory
Iowa State University
6405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
Tel: 515-294-0975
Fax: 515-294-2299
Email: schnable@iastate.edu

Individual basecall and confidence value were assigned using the
Phred software,
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm#b>
rt). Overall sequence quality assessment and vector trimming were
conducted using the Lucy software (<http://www.tigr.org/softlab/>).
Lucy parameters were set to ensure an overall trimmed quality of
97.5% or better without any vector fragments in the chosen
high-quality region of each sequence. Low-quality bases between the
poly-T and the high-quality region were replaced with N's to serve

FEATURES

source

as spacers.
PCR Primers
FORWARD: Forward PCR primer sequence, primer T7-1 (AA TAC GAC TCA
CTA TAG)
BACKWARD: Backward PCR primer sequence, primer T3 (ATT AAC CCT CAC
TAA AG)
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

Location/Qualifiers
1. .644
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="MEST122-F07"
/clone_1lb="ISUM4-TN"
/tissue type="Seedling and silk"
/lab_host="DH10B"

/note="Vector: pT73PAC; Site 1: EcoRI; Site 2: NotI;
ds-cDNA molecules were generated as follows. First-strand
cDNA was prepared from oligo-dT selected mRNA by priming
with a NotI oligo-dT primer (5',
AAGTGAAGATTCGGCGCGAGAAATTTTTTTTTTTTTT). The
resulting DNA:RNA hybrid was treated with RNase H and used
as a template for DNA Pol-I-catalyzed second strand
synthesis. After the addition of EcoRI adaptors, the
ds-cDNAs were digested with NotI and size-selected. The
resulting molecules were directionally cloned into the
EcoRI and NotI sites of the pT73PAC vector. The library
then went through one round of normalization to COT value
of 5 based on the methods of Marcelo Bento Soares (Genome
Research 6: 791-806, 1996)."

BASE COUNT 168 a 154 c 121 g 201 t
ORIGIN

Query Match 71.0%; Score 22; DB 13; Length 644;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;

Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATCCCGCCTTCAATG 31

Db 485 TTACACCTTAACATCCCGCCTTCAATG 514

RESULT 9
AZ495987
LOCUS

DEFINITION 1M0332J11F Mouse 10kb plasmid UUGCM library Mus musculus genomic
clone UUGCM0332J11 F, DNA sequence.

ACCESSION AZ495987
VERSION AZ495987.1 GI:10671822

KEYWORDS
SOURCE
ORGANISM

house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacom, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)

TITLE

JOURNAL

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0332 row: J column: 11
Seq primer: CGTTGTAACGACGCGCAGT

Class: plasmid ends
High quality sequence stop: 670.
Location/Qualifiers

FEATURES

source

1. 670
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCM032311"
/clone_lib="Mouse 10kb plasmid UGCM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42hy; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (g1473214[gblAR129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT
ORIGIN

217 a 124 c 121 g 208 t

Query Match 71.0%; Score 22; DB 17; Length 670;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTACATTCGCCCTTCATG 31
Db 536 TTCACCTTTACTTCTGCTTCCACTG 565

RESULT 10
BM337587 736 bp mRNA linear EST 16-JAN-2002
LOCUS MEST209-B11.T3 ISUMS-RN Zea mays cDNA clone MEST209-B11 3', mRNA
DEFINITION
BM337587 736 bp mRNA linear EST 16-JAN-2002
VERSION
BM337587.1 GI:18167747
KEYWORDS
EST.
SOURCE
Zea mays.
ORGANISM
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
1 (bases 1 to 736)
Men.T.J., Qiu.F., Guo.L., Ashlock.D.A and Schnable.P.S.
Expressed Sequence Tags from B73 Maize: various stages and tissues
including seedlings treated with a variety of hormones
Unpublished (2001)
Contact: Patrick S. Schnable
Schnable Laboratory
Iowa State University
GA05 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
Tel: 515-294-0975
Fax: 515-294-2299
Email: schnable@iastate.edu
Individual basecall and confidence value were assigned using the
phred software,
(<http://depts.washington.edu/ventures/collabr/direct/index.htm#b>
rt). Overall sequence quality assessment and vector trimming were
conducted using the Lucy software (<http://www.tigr.org/softlib/lucy/>).
Lucy parameters were set to ensure an overall trimmed quality of

97.5% or better without any vector fragments in the chosen
high-quality region of each sequence. Low-quality bases between the
poly-T and the high-quality region were replaced with N's to serve
as spacers.

PER PRIMERS
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

FEATURES

source

1. 736
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="MEST209-B11"
/clone_lib="ISUMS-RN"
/tissue_type="mixed"
/lab_host="DH10B"
/note="Vector: pT73PAC; Site 1: EcoRI, Site 2: NotI;
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),
Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels
(3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65 DAG
, Tassel (3-39 cm, 53 and 56 DAG), Silk, unpollinated
cm, 53, 56, 59 DAG), Husk (73 DAG), Silk, unpollinated
first ear, ear shank, etiolated seedlings, callus,
Cycloheximide-treated callus, Anaerobic treated seedlings,
NAA (a-Naphthalene acetic acid)-treated seedlings,
Kineth-treated seedlings, ACP
(1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
Brassinolide-treated seedlings, ABA (Abscissic acid
) -treated seedlings, GA (Gibberellic acid)-treated
seedlings, JA (Jasmonic acid)-treated seedlings, ds-cDNA
molecules were generated as follows. First-strand cDNA was
prepared from oligo-dT selected mRNA by priming with a
NotI oligo-dT primer (5',
AAGTGAAGATTCGCGCGCGCGCAATTTTCTTTTCTTTT). The
resulting DNA:RNA hybrid was treated with RNase H and used
as a template for DNA Pol-catalyzed second strand
synthesis. After the addition of EcoRI adaptors, the
ds-cDNAs were digested with NotI and size-selected. The
resulting molecules were directionally cloned into the
EcoRI and NotI sites of the pT73PAC vector. The library
then went through one round of normalization to Cot value
of 5 based on the methods of Marcelo Bento Soares (Genome
Research 6: 791-806, 1996)."

BASE COUNT 181 a 164 c 146 g 225 t 20 others

ORIGIN

Query Match 71.0%; Score 22; DB 13; Length 736;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTACATTCGCCCTTCATG 31
Db 538 TTCACCTTTACATTCGCCCTTCATG 567

RESULT 11
AY108329/c 920 bp mRNA linear HTC 25-MAY-2002
LOCUS AY108329 920 bp mRNA linear HTC 25-MAY-2002
DEFINITION
AY108329 920 bp mRNA linear HTC 25-MAY-2002
VERSION
AY108329.1 GI:21211407
KEYWORDS
HTC.
SOURCE
Zea mays.
ORGANISM
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
1 (bases 1 to 920)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitelc,M.S.,
Athur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/Dupont Consensus Sequences for Design of
Overgo Probes

JOURNAL Unpublished (2002)
 REFERENCE 2 (bases 1 to 920)
 AUTHORS Coe,B.C.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 FEATURES
 source
 1. 920
 /organism="Zea mays"
 /db_xref="taxon:6239"
 /db_xref="taxon:4577"
 /clone_lib="PCO128508"
 /note="Maize Mapping Project/DuPont Cornsensus
 library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

BASE COUNT 230 a 247 c 224 g 213 t 6 others
 ORIGIN

Query Match 71.0%; Score 22; DB 11; Length 920;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCCTTTCATG 31
 Db 763 TTACACCTTTAAATTCGCCCTTTCATG 734

RESULT 12
 LOCUS M79865 384 bp mRNA linear EST 30-JUN-1992
 DEFINITION WIST00402 Mixed stage, Stratiagene (cat. #937006) Caenorhabditis
 elegans CDNA clone CEMSE75, mRNA sequence.
 M79865
 M79865.1 GI:271884
 EST.
 Caenorhabditis elegans.
 Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea
 ; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 384)
 McCombie,W.R., Adams,M.D., Kelley,J.M., Fitzgerald,M.G., Uterback
 T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J. and Fields,C.
 Caenorhabditis elegans expressed sequence tags identify gene
 families and potential disease gene homologues
 Nat. Genet. 1, 124-131 (1992)
 JOURNAL 93250983
 MEDLINE
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 Seq primer: M13 Forward.
 Location/Qualifiers
 1. 384
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone_lib="CEMSE75"
 /note="Mixed stage, Stratiagene (cat. #937006)"
 /note="Vector: Uni-ZAP XR; C. elegans mixed stage CDNA
 library. Stratiagene catalog #937006. The library is oligo
 dt primed and directionally cloned in the Uni-ZAP XR
 vector."

BASE COUNT 156 a 46 c 74 g 101 t 7 others
 ORIGIN

Query Match 69.7%; Score 21.6; DB 14; Length 384;

Best Local Similarity 80.0%; Pred. No. 2.1e+02;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATTCCACCTTTTACATTCGCCCTTTCAT 30
 Db 338 ATTCCACCTTTTAAATTCGCCCTTTCAT 309

RESULT 13
 LOCUS AO500362 605 bp DNA linear GSS 29-APR-1999
 DEFINITION Y41F1 m7n-3xHA/lacZ insertion library Saccharomyces cerevisiae
 genomic 5', DNA sequence.
 AO500362
 AO500362.1 GI:4706072
 GSS.
 baker's yeast.
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 1 (bases 1 to 605)
 Rose-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A.,
 deBerges,S.A., Cheung,K.-H., Sheehan,A., Symonakis,D., Hansen,R.,
 Umaneky,L., Heidman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R.,
 Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
 Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
 Gene Disruption
 Unpublished (1999)
 Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumare@yale.edu
 Seq primer: GGCCTTCCTTCTTTCGACAGTAC
 Class: transposon-tagged.
 Location/Qualifiers
 1. 605
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /clone_lib="m7n-3xHA/lacZ insertion library"
 /lab_host="S. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library
 (lacking mitochondrial DNA) was prepared in pHS6-Sal;
 genomic DNA was size-fractionated (DNA of roughly 2-3 kb
 in length) prior to cloning. This library was
 subsequently mutagenized with a m7n-3xHA/lacZ
 minitransposon containing lacZ, URA3, and tet resistance."

FEATURES
 source
 1. 605
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /clone_lib="m7n-3xHA/lacZ insertion library"
 /lab_host="S. coli"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library
 (lacking mitochondrial DNA) was prepared in pHS6-Sal;
 genomic DNA was size-fractionated (DNA of roughly 2-3 kb
 in length) prior to cloning. This library was
 subsequently mutagenized with a m7n-3xHA/lacZ
 minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 160 a 149 c 108 g 188 t
 ORIGIN

Query Match 69.7%; Score 21.6; DB 17; Length 605;
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TTCACCTTTTACATTCGCCCTTTCAT 29
 Db 578 TTCACCTTTTAAATTCGCCCTTTCAT 605

RESULT 14
 LOCUS BG355209 447 bp mRNA linear EST 06-MAR-2001
 DEFINITION 947040C05.y2 947 - 2 week shoot from Barikan lab Zea mays CDNA, mRNA
 sequence.
 BG355209
 BG355209.1 GI:13237195
 EST.
 Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Zea mays.
 SOURCE
 ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 447)
Walbot, V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 947040 row: C column: 05.
Location/Qualifiers

FEATURES
SOURCE

1..447
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="947 - 2 week shoot from Barkan lab"
/issue_type="leaf and stem, including leaf base"
/dev_stage="2 week old seedling (3 leaves)"
/lab_host="XLI-Blue"
/note="Organ: shoot; Vector: Lambda ZAP (pbluescript SK-);
Site 1: ECORI; Site 2: XhoI; Directionally cloned using
Stratagene's Unitap XR cDNA cloning kit with the 5' end
at the ECORI site. The library represents 8 x 10⁵
independent recombinant phage. The plants were greenhouse
grown."

BASE COUNT 137 a 78 c 112 g 120 t
ORIGIN

Query Match 69.0%; Score 21.4; DB 12; Length 447;
Best Local Similarity 80.6%; Pred. No. 2.6e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ATTCACCTTTACATTCGCCCTTCATTG 31
Db 334 ATTCTCCTTTACTTCTTGCTTCCTTG 304

RESULT 15
AM331076 477 bp mRNA linear EST 31-JAN-2000
LOCUS 707047A01.x1 707 - Mixed adult tissues from Walbot lab (SK) Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION AM331076
VERSION AM331076.1 GI:6827433
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 477)
Walbot, V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 707047 row: A column: 01.
Location/Qualifiers

FEATURES
SOURCE

1..477
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"

/clone_lib="707 - Mixed adult tissues from Walbot lab (SK
)"
/issue_type="tassel, kernel, silk, husk, root, leaf"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: tassel, kernel, silk, husk, root, leaf;
Vector: pGAD10; Site 1: ECORI; cDNA library from fully
differentiated maize tissues from an active Mutator
plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk,
husk, root, leaf). Unidirectionally cloned."

BASE COUNT 155 a 70 c 116 g 136 t
ORIGIN

Query Match 69.0%; Score 21.4; DB 10; Length 477;
Best Local Similarity 80.6%; Pred. No. 2.7e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ATTCACCTTTACATTCGCCCTTCATTG 31
Db 163 ATTCTCCTTTACTTCTTGCTTCCTTG 133

Search completed: May 11, 2003, 04:54:34
Job time : 502.331 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 338.927 Seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-18

Perfect score: 32

Sequence: 1 caagggctcgaataaagaataatcgtga 32

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : *

1: gb Da:*
2: gb Htg:*
3: gb In:*
4: gb Om:*
5: gb Ov:*
6: gb Pac:*
7: gb Ph:*
8: gb Pl:*
9: gb Pr:*
10: gb Ro:*
11: gb Sts:*
12: gb Sy:*
13: gb Un:*
14: gb Vi:*
15: em Ba:*
16: em Fun:*
17: em Hum:*
18: em In:*
19: em Mu:*
20: em Om:*
21: em Or:*
22: em Ov:*
23: em Pat:*
24: em Ph:*
25: em Pl:*
26: em Ro:*
27: em Sts:*
28: em Un:*
29: em Vi:*
30: em Htg_Hum:*
31: em Htg_Inv:*
32: em Htg_Other:*
33: em Htg_Mus:*
34: em Htg_Pln:*
35: em Htg_Rod:*
36: em Htg_Mam:*
37: em Htg_Vrc:*
38: em Sy:*
39: em Htgo_Hum:*
40: em Htgo_Mus:*
41: em Htgo_Other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	32	100.0	32	6	AX036752	AX036752 Sequence
2	22.4	70.0	116498	2	OSJ000004	AL064449 Oryza sat
3	22.4	70.0	167191	2	AC123085	AC123085 Rattus no
4	22.4	70.0	208214	10	AL645908	AL645908 Mouse DNA
5	22	68.8	213383	10	AL645943	AL645943 Mouse DNA
6	21.4	66.9	41502	3	CEZK856	270783 Caenorhabdi
7	21.4	66.9	158886	2	AC016944	AC016944 Homo sapi
8	21.4	66.9	159981	9	AL161899	AL161899 Human DNA
9	21.4	66.9	188215	2	AC025725	AC025725 Caenorhab
10	21.4	66.9	188616	3	AC084158	AC084158 Caenorhab
11	21.4	66.9	206320	9	AC093418	AC093418 Homo sapi
12	21.4	66.9	229081	2	AC006892	AC006892 Caenorhab
13	21.2	66.2	1567	6	AR016866	AR016866 Sequence
14	21.2	66.2	1567	6	AR020892	AR020892 Sequence
15	21.2	66.2	1567	6	AR027215	AR027215 Sequence
16	21.2	66.2	1567	6	AR038502	AR038502 Sequence
17	21.2	66.2	1567	6	AR064644	AR064644 Sequence
18	21.2	66.2	1567	6	AR067569	AR067569 Sequence
19	21.2	66.2	1567	6	I38521	I38521 Sequence 10
20	21.2	66.2	1567	6	I56996	I56996 Sequence 10
21	21.2	66.2	1567	6	I59862	I59862 Sequence 10
22	21.2	66.2	1567	6	I75189	I75189 Sequence 10
23	21.2	66.2	94727	9	AL591034	AL591034 Human DNA
24	21.2	66.2	114468	9	AC093026	AC093026 Homo sapi
25	21.2	66.2	162808	2	AC021340	AC021340 Homo sapi
26	21.2	66.2	164528	2	AC016800	AC016800 Homo sapi
27	21.2	66.2	169161	2	AC019053	AC019053 Homo sapi
28	21.2	66.2	171490	2	AC068209	AC068209 Homo sapi
29	21	65.6	2898	6	AX010976	AX010976 Sequence
30	21	65.6	2898	6	PFU278385	PFU278385 Penicilli
31	21	65.6	229061	9	AC007786	AC007786 Homo sapi
32	20.8	65.0	3103	1	AF269420	AF269420 Staphyloc
33	20.8	65.0	3103	6	AX144740	AX144740 Sequence
34	20.8	65.0	3438	1	AF270201	AF270201 Staphyloc
35	20.8	65.0	3438	6	AX145519	AX145519 Sequence
36	20.8	65.0	3463	1	AF269557	AF269557 Staphyloc
37	20.8	65.0	3463	6	AX145275	AX145275 Sequence
38	20.8	65.0	38621	2	AC079778	AC079778 Homo sapi
39	20.8	65.0	55835	2	AL138772	AL138772 Human DNA
40	20.8	65.0	100886	2	AP003831	AP003831 Oryza sat
41	20.8	65.0	111500	9	AC121344	AC121344 Homo sapi
42	20.8	65.0	113071	2	CNS07YPK	AL713938 Oryza sat
43	20.8	65.0	115393	8	AC091247	AC091247 Oryza sat
44	20.8	65.0	117505	8	AC096687	AC096687 Oryza sat
45	20.8	65.0	122381	9	AP000577	AP000577 Homo sapi

ALIGNMENTS

RESULT 1
AX036752
LOCUS
DEFINITION Sequence 18 from Patent WO0058485.
ACCESSION AX036752
VERSION AX036752.1 GI:11226261
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 32)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 18 05-OCT-2000

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
FEATURES
Location/Qualifiers
1. 32
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Guide desoxynucleotide building block G4"

BASE COUNT 14 a 3 c 8 g 7 t
ORIGIN

Query Match 100.0%; Score 32; DB 6; Length 32;
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAAGGTCGAATTAAGAAATTAATTCGTGA 32
Db 1 CAAGGTCGAATTAAGAAATTAATTCGTGA 32

RESULT 2
OSJN00004/c 116498 bp DNA linear HTG 17-JUL-2002
DEFINITION Oryza sativa chromosome 4 clone OSJNBA0014F04, *** SEQUENCING IN
PROGRESS ***, 2 ordered pieces.
ACCESSION AL606449
VERSION AL606449.2 GI:21912451
KEYWORDS HTG; HTGS PHASE1
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Euphorbiaceae; Oryzaceae; Oryza.

REFERENCE

1 Pu, G., Wang, S. Y., Ren, S. X., Lv, G., Lin, W., Gu, W. Q., Zhu, G. F.,
Tu, Y. F., Jia, J., Yin, H. P., Zhang, Y., Cai, Z., Chen, J., Kang, H.,
Chen, X. Y., Shao, C. Y., Sun, Y., Hu, Q. P., Zhang, X. L., Zhang, W.,
Wang, L. J., Ding, C. W., Sheng, H. H., Gu, J. L., Chen, S. T., Ni, L.,
Zhu, F. H., Han, B., Feng, Q., Huang, Y. C., Li, Y., Zhu, J. J., Zhao, Q.,
Hu, X., Liu, Y. L., Mu, J., Yu, Z., Chen, L., Fan, D. L., Meng, Q. J.,
Zhang, L., Lu, Y. Q., Yu, S. L., Liu, X. H., Lu, T. T., Zhang, Y. Q., Lu, Y.,
Li, C., Li, T., Zhang, Y., Hu, H., Jia, P. X., Qian, Y. M., Ying, K.,
Zhou, B., Chen, Z. H., Hao, P., Zhang, L., Wu, M., Zhang, R. Q., Guan, J. P.
and Hong, G. P.
Direct Submission
Submitted (27-JUN-2002) Han Bin, National Center for Gene Research,
Chinese Academy of Sciences, 500# Cao Bao Road, Shanghai 200233,
CHINA. E-mail enquiries: dhancngr.ac.cn. Clone requests:
dhancngr.ac.cn
Oryza sativa Japonica (nipponbare) genomic DNA, chromosome 4, BAC
clone: OSJNBA0014F04.
On Jul 19, 2002 this sequence version replaced gi:15552687.
Web site: http://www.ncgr.ac.cn

TITLE

JOURNAL

REMARK

COMMENT

Assembly program: Phrap
Summary Statistics

NOTE: This is a PHASE2 sequence. Gaps are shown by nnnn. Genes
were identified by a combination of several methods: Gene
prediction programs including Rgenesh (http://www.socberry.com/),
genescan (http://CCR-081.mit.edu/GENSCAN.html), GeneMarkES
(http://genemark.biology.gatech.edu/GeneMark/), CRNAcan-SE (Sean
Eddy, http://genome.msucl.edu/eddy/CRNAcan-SE/), searches of the
complete sequence against NCBI non-redundant protein database (nr)
(ftp://ncbi.nlm.nih.gov/blast/db) and the EST database at NCGR.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 62574: contig of 62574 bp in length

* 62575 62774: gap of 200 bp
* 62775 116498: contig of 53724 bp in length.
FEATURES
Location/Qualifiers
1. 116498
/organism="Oryza sativa"
/variety="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/chromosome="4"
/clone="OSJNBA0014F04"
/clone_1ib="CUGI-OSJNBA"

BASE COUNT 32679 a 26432 c 26711 g 30475 t 201 others
ORIGIN

Query Match 70.0%; Score 22.4; DB 2; Length 116498;
Best Local Similarity 81.2%; Pred. No. 1.5e+02;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 CAAGGTCGAATTAAGAAATTAATTCGTGA 32
Db 13070 CAAGATTGCACTTGGGAAACAAATTCGTGA 13039

RESULT 3
AC123085
LOCUS
DEFINITION Rattus norvegicus clone CH230-909, *** SEQUENCING IN PROGRESS ***
ACCESSION AC123085
VERSION AC123085.2 GI:21723033
KEYWORDS HTG; HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 167191)
Alzobrooke, S. L., Amaral, C., Adio-Oduola, B., Ali-Adnan, F. R., Allen, C.,
Mazny, D. M., Adams, C., Aze, J. R., Ayale, M., Banks, T.,
Barbieri, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,
Buck, J., Bowie, S., Brice, M., Brown, E., Brown, M., Bryant, N. P.,
Buhay, C., Burich, P., Burkett, C., Burrell, K. L., Byrd, N. C.,
Carton, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C. D., Cox, C., Coyle, M. D., Dathorne, S. R., David, R.,
Davila, M. L., Davis, C., Davy-Carroll, L., Dedrich, D. A.,
Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H.,
Douthwaite, K. J., Draper, H., Dugan-Rocha, S., Durbin, K. J.,
Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M.,
Falle, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabriel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorelli, J. H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
Homes, F., Howard, S., Huber, J., Huijck, S., Hume, J., Jackson, L. E.,
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C.,
Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,
Li, J., Li, Z., Lichner, O., Lien, C., Liu, J., Liu, W., Louie, H.,
Lozadhar, M., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabac, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, E., Nwokenkwo, S., Ogih, M., Okwunonu, G.,
Oguyun, N., Nicholson, E., Nwokenkwo, S., Ogih, M., Okwunonu, G.,
Oguyun, N., Oyler, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pichens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojupokan, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sleson, I.,
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansley, J., Taylor, C., Taylor, T., Talford, B., Thomas, N., Thomas, S.,
Umami, K., Vaequez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 167191)
Worley, K.C.
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
(bases 1 to 167191)
Worley, K.C.
Direct Submission
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 10, 2002 this sequence version replaced gi:21239777.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GDXP
Center clone name: CH230-909
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 114513 bases at least Q40
Consensus quality: 117131 bases at least Q30
Consensus quality: 119267 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1066: contig of 1066 bp in length
1067
1166: gap of unknown length
1167
2757: contig of 1591 bp in length
2758
2857: gap of unknown length
2859
4142: contig of 1285 bp in length
4143
4242: gap of unknown length
4243
5847: contig of 1605 bp in length
5848
5947: gap of unknown length
5948
7339: contig of 1392 bp in length
7340
7439: gap of unknown length
7440
8655: contig of 1216 bp in length
8656
8755: gap of unknown length
8756
10233: contig of 1478 bp in length
10234
10333: gap of unknown length
10334
12015: contig of 1682 bp in length
12016
12115: gap of unknown length
12116
13329: contig of 1114 bp in length
13329
13329: gap of unknown length
13330
15087: contig of 1758 bp in length
15087
15187: gap of unknown length
15188
15876: contig of 1589 bp in length
15877
16777: gap of unknown length
16778
18703: contig of 1827 bp in length
18704
18803: gap of unknown length
20293: contig of 1490 bp in length
20294
20393: gap of unknown length
20394
21718: contig of 1325 bp in length
21719
21818: gap of unknown length

21819
23075
23175
24806
24906
24906
26255: contig of 1350 bp in length
26256
26356
27516
27615
28892: contig of 1277 bp in length
28893
28993
30170: contig of 1178 bp in length
30171
30270: gap of unknown length
30271
31632: contig of 1362 bp in length
31633
31732: gap of unknown length
31733
33023: contig of 1291 bp in length
33024
33123
33123: gap of unknown length
33124
34124: contig of 1001 bp in length
34125
34224: gap of unknown length
34225
35778: contig of 1554 bp in length
35779
35878: gap of unknown length
35879
37834: contig of 1956 bp in length
37835
37934: gap of unknown length
37935
40210: contig of 2276 bp in length
40211
40310: gap of unknown length
40311
41638: contig of 1328 bp in length
41738
41738: gap of unknown length
43521: contig of 1783 bp in length
43522
43621: gap of unknown length
43622
45077: contig of 1456 bp in length
45078
45177: gap of unknown length
45178
46984: contig of 1807 bp in length
47084
47084: gap of unknown length
47085
48843: contig of 1759 bp in length
48844
48943: gap of unknown length
48944
50980: contig of 2037 bp in length
50981
51080: gap of unknown length
52334: contig of 1254 bp in length
52335
52434: gap of unknown length
54389: contig of 1955 bp in length
54390
54490
56000: gap of unknown length
56001
56100: gap of unknown length
56101
58503: contig of 2403 bp in length
58504
58603: gap of unknown length
58604
60246: contig of 1643 bp in length
60247
60346: gap of unknown length
60347
60347: contig of 3087 bp in length
63433
63433: gap of unknown length
63434
65154: contig of 2621 bp in length
65155
66254: gap of unknown length
66255
67965: contig of 1711 bp in length
67966
68065: gap of unknown length
68066
70031: contig of 1966 bp in length
70032
70131: gap of unknown length
72594
72594: contig of 2463 bp in length
72595
72694: gap of unknown length
75301: contig of 2607 bp in length
75302
75401: gap of unknown length
75402
79068: contig of 3667 bp in length
79069
79168: gap of unknown length
80614
80614: contig of 1446 bp in length
80714
80714: gap of unknown length
83675
83675: contig of 2961 bp in length
83676
83775: gap of unknown length
83776
86937: contig of 3162 bp in length
86938
87037: gap of unknown length
90098
90098: contig of 3061 bp in length
90198
90198: gap of unknown length
92410
92410: contig of 2212 bp in length
92411
92510: gap of unknown length
92511
94924: contig of 2414 bp in length
94925
95024: gap of unknown length
100832: contig of 5808 bp in length

* 100833 100932: gap of unknown length
 * 100933 104553: contig of 3621 bp in length
 * 104554 104653: gap of unknown length
 * 104654 108005: contig of 3352 bp in length
 * 108006 108105: gap of unknown length
 * 108106 112900: contig of 4795 bp in length

Query Match 70.0% Score 22.4; DB 2; Length 167191;
 Best Local Similarity 81.2%; Pred. No. 1.4e+02;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CAAGGTTGGAATTAAGCAATTAATTCGTGA 32

Db 99808 CAAGGTTGGAATTAAGCAATTAATTCGTGA 99839

RESULT 4

AL645908/c 208214 bp DNA linear ROD 05-APR-2002
 LOCUS Mouse DNA sequence from clone RP23-24A24 on chromosome 11, complete
 DEFINITION sequence.

ACCESSION AL645908 GI:20068640

VERSION AL645908.9

KEYWORDS HTG.

SOURCE house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Pearce, A.
 Direct Submission
 Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Apr 7, 2002 this sequence version replaced gi:117976640.

COMMENT

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SwissProt; Tr: TrEMBL; Wp: WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-24A24 is from the RP23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6.

FEATURES

Source 1..208214

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-24A24"
 /clone_1lb="RP23-23"

BASE COUNT 64160 a 43875 c 41691 g 58484 t

ORIGIN

Query Match 70.0% Score 22.4; DB 10; Length 208214;
 Best Local Similarity 81.2%; Pred. No. 1.4e+02;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 CAAGGTTGGAATTAAGCAATTAATTCGTGA 32

||||| ||||| ||||| ||||| |||||

Db 89603 CAAGTGTGGAATTAAGAGTAAGCA 89572

RESULT 5

AL645943/c 213383 bp DNA linear ROD 12-JUL-2002
 LOCUS Mouse DNA sequence from clone RP23-130E22 on chromosome 11,
 DEFINITION complete sequence.

ACCESSION AL645943 GI:22138910

VERSION AL645943.19

KEYWORDS HTG.

SOURCE

house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 213383)

REFERENCE

1 Clark, S.
 Direct Submission
 Submitted (12-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Aug 8, 2002 this sequence version replaced gi:21732076.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SwissProt; Tr: TrEMBL; Wp: WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-130E22 is from the RP23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6.

FEATURES

Source 1..213383

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /clone="RP23-130E22"
 /clone_1lb="RP23-23"

BASE COUNT 59658 a 48025 c 47605 g 58091 t

ORIGIN

Query Match 68.8% Score 22; DB 10; Length 213383;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 25; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 AAGGTTGGAATTAAGCAATTAATTCGTG 31

Db 70075 AAGGTTGGAATTAAGCAATTAATTCATG 70046

RESULT 6

CEZK856 41502 bp DNA linear INV 24-JAN-2002
 LOCUS Caenorhabditis elegans cosmid ZK856, complete sequence.
 DEFINITION

ACCESSION 270783
VERSION 270783.1 GI:1263486
KEYWORDS HTG; DNA directed RNA polymerase III; EF-hand; LIN-19; Spingomyelin phosphodiesterase; TAU; Transposase.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1
AUTHORS none.
TITLE Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
PUBMED 9851916
REMARK The *C. elegans* Sequencing Consortium.
REFERENCE 2 (bases 1 to 41502)
AUTHORS Lloyd, C.R.
JOURNAL Direct Submission
TITLE Submitted (16-APR-1996) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jess@anger.ac.uk or r.wenematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.
Current sequence finishing criteria for the *C. elegans* genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone ZK856. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone ZK856 is at 32458 in sequence Z78410.
The true right end of clone ZK856 is at 104 in sequence Z73969.
The start of this sequence (1..104) overlaps with the end of sequence Z78410.
The end of this sequence (41399..41502) overlaps with the start of sequence Z73969.
For a graphical representation of this sequence and its analysis see: <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=ZK856>
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
location/Qualifiers
1..41502
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"
/clone="ZK856"
join(2797..2859,2905..3039,3418..3519,3567..3771,3818..3927,3975..4182,4229..4359,4409..4705,4752,4995..5055,5102..5205,52370..5534,5637..6028,6074..6125)/gene="ZK856.1"
join(2797..2859,2905..3039,3418..3519,3567..3771,3818..3927,3975..4182,4229..4359,4409..4705,4752,4995..5055,5102..5205,5370..5534,5637..6028,6074..6125)/gene="ZK856.1"
/note="Similarity with the *C. elegans* LIN-19 protein, contains similarity to Pfam domain: PF00888 (Cullin family), Score=1185.1, E-value=0, N=1
CDNA EST YK5993.3 comes from this gene
CDNA EST YK5993.5 comes from this gene
CDNA EST YK16197.5 comes from this gene
CDNA EST YK16197.3 comes from this gene
CDNA EST YK4623.3 comes from this gene
CDNA EST YK4623.5 comes from this gene

CDNA EST YK310H9.3 comes from this gene
CDNA EST YK34897.3 comes from this gene
CDNA EST YK34897.5 comes from this gene
CDNA EST YK36758.5 comes from this gene
CDNA EST YK539411.3 comes from this gene
CDNA EST YK539411.3 comes from this gene
CDNA EST YK570C12.3 comes from this gene
CDNA EST YK539411.5 comes from this gene
CDNA EST YK570C12.5 comes from this gene
/codon_start=1
/protein_id="CA94852.1"
/db_xref="GI:3881806"
/translation="MQDEWSKADPIVHALHQKSVTPAAMODLFYHYVTKTSWDD
GPKLIDILTRCINDYVHANKRIRSLQDGSILYIKEMNFPYQOANLIPFKIK
DESSRRSVPEPEESIRVTMLEKNEIIFMNSISQLEALYEAERNGNIDN
VIGIRSFVALNDRAGEDPLVYROSPEEQIFQETBYKKICGNLNLGIVLEVY
ADKLEEQORAKRYLENNPSTGKMEKAVIALVSFEDTILAECSKILASQVETL
QRLYLIRRTSGIDTFLKCIDTHIRKTEGLMDNRNNAENLSTPERYVOQLAMPDE
SLVREGFCDNALTLTARDKAFRAVAVNDISIPTENMNNKGRITLSVESKCAELIAYC
DILLRTQLSKLTSEIDEKLNQVLLVLYKENKQVFMFRFAHLSRLILMSADQ
EKEMWVTLRREGCMPSDAVNKLRLQDIELKNNSSPKKALGTNNKNSIADSN
MYLNGANGRGSSERIRPSLPREDPVPDEWPKRANGRPDLCTTROMALPCGN
DRANDKISLETALATLTPDAELNRTLSLVAYPKRQIILCDVSTTYTARDTDS
TKEFLNHDNNAVNGKSOQGRKNLLIGRLQLSLEANAERKHSIVALRELVOEGIV
ILKRTKTYTLAQITMELVILKPLFIPNRKIKKEQIDWLIENKYMERRADINTVYI
S"
join(6516..6596,7028..7111,7164..7379,7429..7540,7584..7678)
/gene="ZK856.4"
join(6516..6596,7028..7111,7164..7379,7429..7540,7584..7678)
/gene="ZK856.4"
/codon_start=1
/protein_id="CA94851.1"
/db_xref="GI:3881805"
/translation="MFEHQLLKMKLYIILCTIVASEALDFNPLHQLCWLQVAVILA
KRTNSQAVKIEIKTCNNVLLDNNWVCNGLIEVSAFPGTTPAVKLCNLMYCASS
DKRSKMSKPVKGRPNELTKVYLDLLENDEPVSQDRFRQVAVNEALINFGN
DICKMLQGVTFPNNDIYHFLKKPQKQISVSGENE"
complement(join(7983..8215,8261..8426,8632..8932,8984..9175,9380..9520,9590..9715,9760..9852,9897..10179,10239..10410))
/gene="ZK856.3"
complement(join(7983..8215,8261..8426,8632..8932,8984..9175,9380..9520,9590..9715,9760..9852,9897..10179,10239..10410))
/gene="ZK856.3"
/note="Weak similarity with spingomyelin phosphodiesterase"
/codon_start=1
/protein_id="CA94850.1"
/db_xref="GI:3881804"
/translation="MNSFSANKAYVNFYTKKVPADQFAFLKELSNAKCPRNSENC
TSIVLVAHIGPEVPEKTPNFMFRDENRPFDLIVRVANSIGMMIFGHHDTFL
INDSKENNQALMAPVTPWPSDGLGACANNTPFRVETDASIKODISTYINND
LNNKNSIPVPEFSFKDAIGITGIDINPTTMSALDPMKNNDTLFFKTIIDNTAFMPO
MPLGVYRGALQSMESYSDYPRYSSCLSTYSALTRSLNLPFLFPAIITYELALFQ
SKVHLHFOLEFSDIHTRKNSLYOHPKKKESKPNALILNLTGTFGEKFEVDVQBEL
APGGLIYAVOICPIDACQAVVQNDYERSARQVRLLELTGTFVNGVMYTCIP
VKNHVARTHREEREVEPGLTTSRGVTLVDSGIFASFTQONSVSQTSIFSTOSF
SSYSLSRSRRTSSSSQYABSDSSATITTPPSYTAQOSTGCLNLPPTHTYPTDAF
FRRTKILIPPEMKQELAPVTOQIMNCYINSIRKGSDBIGRQSLSCSKMELBYNWTG
SL"
complement(join(11155..12030,12072..12227))
/gene="ZK856.2"
complement(join(11155..12030,12072..12227))
/gene="ZK856.2"
/note="contains similarity to Pfam domain: PP01498 (Transposase), Score=208.2, E-value=4.1e-59, N=1"
/codon_start=1
/protein_id="CA94849.1"

```

/db_xref="GI:3881803"
/translation="MVKSGCKNLSLVKAKIYVGFQGIPTKMLALQIORSPTIK
VIRKRRKQVALRISPRPRVTHRMDDRIISAPRDPRITDIOITSSPREPE
SKTVRRRLQAGLHGRKPYKPPFISCKRMAYAKALHNRGOSMAHINSDESK
FNLFSGSNGSWRRPVGSRYSPTQCTVHGGGVVWVWCFTSFGPLRRIOSID
RKYENITLETTRMVALQNVGRGVFOQDDPKHTSLATVSWFORRRVHLIDWPSOP
DLNPIEHLMEELERLLGIGIRASNAKAFENLNAWKAI PMSVIHLIDSMPRCQAVI
DANGVATKY"
complement(join(12949..13101,13147..13350))
/gene="ZK856.5"
complement(join(12949..13101,13147..13350))
/gene="ZK856.5"
/codon_start=1
/codon_id="CAA94853.2"
/db_xref="GI:18376575"
/db_xref="SPTREMBL:Q23640"
/translation="MFMNLMYLLFLPLPKSGWQVYLHADPLVDVYSINGDNKRM
CHDDGRNTLAGEYGDYCDAPKPLQLAIDBSARLPNPDLLIWTGDNVAHIDYIG
WESKLFSCIOYVSTL"
join(21936..21996,22068..22271,22325..22439,22490..22583,
22883..22945)
/gene="ZK856.6"
join(21936..21996,22068..22271,22325..22439,22490..22583,
22883..22945)
/gene="ZK856.6"
/codon_start=1
/codon_id="CAA94854.1"
/db_xref="GI:3881808"
/db_xref="SPTREMBL:Q23641"
/translation="MASVFLMLVLTITIAVCTYVVVEPLVESMBEAAVALDPE
VPVAPSDPILLRKRSILVLSEKTEVCKTKRQHHRPOYGIANGSHVEIQODE
SSLSEATFECTGEMRPECVGHVDLFISECVTYEHRPAMVAVVSGSPFPAITRV
ILCECRLRROFKDPEKRX"
join(25331..25464,25511..25742)
/gene="ZK856.7"
join(25331..25464,25511..25742)
/gene="ZK856.7"
/codon_start=1
/protein_id="CAA94855.1"
/db_xref="GI:3881809"
/db_xref="SPTREMBL:Q23642"
/translation="MRVAVILFPLSVVALACDIIHLKSDTDKKSQIVASNGKSE
RMSYLKLLQKNTFOQKADCEGLKDMWBITTDDGKPAVATKATLIDGRTYKVDL
KAVOKDROAGIIGSCAPL"
complement(join(25949..26187,26233..26457,26508..26564,
26642..26708))
/gene="ZK856.8"
complement(join(25949..26187,26233..26457,26508..26564,
26642..26708))
/gene="ZK856.8"
/notes="contains similarity to Pfam domain: PF00036 (BF
Query Match
Best Local Similarity 66.9%; Score 21.4; DB 3; Length 41502;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
OY 2 AAGGCTCGAATAAGGAATAATTCGTGA 32
Db 22923 AAGATTTCGAAAAAGGAATAATTCGTGA 22953

```

```

REFERENCE
AUTHORS
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 158886)
Muzny, D.M., Adams, C., Bailey, M., Barbaria, J., Blankenburg, K.,
Bodoca, B., Bouck, J., Bowie, S., Brooks, A., Bunay, C., Bunac, C.,
Burt, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Raahid, N.,
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
Forcum-Tansey, J., Frantz, P., Ganesht, R., Gorrell, J.H., Gorrell, L.L.,
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hogue, M.,
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,
Kelly, S., Kondrjewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
Lichter, O., Liu, J., Liu, W., Logan, O., Lozada, R.J., Lu, J.,
Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mel, G., Morgan, M.,
Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,
Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
Shah, B., Shen, H., Simon, M., Sparks, A., Stamp, A., Sungang, R.,
Taber, P., Taylor, T., Vaequez, L., Vinson, R., Vo, Q., Wahbah, M.,
Wallington, S., Weinstein, G., Weinstein, I.R., Williamson, A.,
Worley, K., Wren, J., Wrenford, G., Yu, W., Zhou, X., Nelson, D. and
Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 158886)
Mozley, K.C.
Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 26, 2000 this sequence version replaced gi:9719630.
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: RP11-70P20
Center clone name: RP11-70P20
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 144562 bases at least Q40
Consensus quality: 152026 bases at least Q30
Consensus quality: 154719 bases at least Q20
Estimated insert size: 155752; sum-of-coverage
Estimated insert size: 209830; agarose-1p estimation
Quality coverage: 3.2x in Q20 bases; agarose-1p estimation
Quality coverage: 4.3x in Q20 bases; sum-of-coverage
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
43752: contig of 43752 bp in length
43753 43852: gap of unknown length
43853 71719: contig of 27867 bp in length
71720 71819: gap of unknown length
71820 93367: contig of 21548 bp in length
93368 93467: gap of unknown length
93468 108864: contig of 15397 bp in length
108865 108964: gap of unknown length
108965 124124: contig of 15160 bp in length
124125 124224: gap of unknown length
124225 135573: contig of 11349 bp in length
135574 135673: gap of unknown length
135674 141767: contig of 6094 bp in length

```

141768 141867: gap of unknown length
* 141868 149244: contig of 7377 bp in length
* 149245 149344: gap of unknown length
* 149345 155749: contig of 6405 bp in length
* 155750 155849: gap of unknown length
* 155850 157240: contig of 1391 bp in length
* 157241 157341: gap of unknown length
* 157341 158886: contig of 1546 bp in length.
Location/Qualifiers
1. 158886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-70P20"

BASE COUNT 45757 a 33003 c 33199 g 45917 t 1010 others

ORIGIN

Query Match 66.9%; Score 21.4; DB 2; Length 15886;
Best Local Similarity 80.6%; Pred. No. 3.4e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AAGGTTGGAATTAAGGATTAATTCGTA 32
Db 105743 AAGGATTGAATTAAGGATTAATTCGTA 105713

RESULT 8
AL161899 159981 bp DNA linear PRI 25-FEB-2001
LOCUS Human DNA sequence from clone RP11-418110 on chromosome 13,
DEFINITION complete sequence.
ACCESSION AL161899
VERSION AL161899.21 GI:13606073
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 159981)
Phillimore,B.
Direct Submission
Submitted (24-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
request: clonerequest@sanger.ac.uk
On Apr 12, 2001 this sequence version replaced gi:12964402.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-418110 is from the library RPCT-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-418110 The true
left end of clone RP11-347N5 is at 72478 in this sequence. The true

right end of clone RP11-307B20 is at 55792 in this sequence.
FEATURES
source
1. 159981
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-418110"
/clone_lib="RPCT-11.2"
26800..26832
/note="Sequence from overlapping clone BA307B20
(AL139802). Assembly confirmed by restriction digest data."
36723..36895
/note="Sequence from overlapping clone BA307B20
(AL139802). Assembly confirmed by restriction digest."
37359..37715
/note="Sequence from overlapping clone BA307B20
(AL139802). Assembly confirmed by restriction digest."
40660..41106
/note="Single clone region. Sequence from reads from a
short insert library derived from a clone PCR. Restriction
digest data confirm the assembly."
67000..67044
/note="Sequence confirmed by overlapping clone AC021003
sequence by Whitehead Institute. Assembly confirmed by
restriction digest."
67045..67080
/note="Sequence from overlapping clone AC021003 sequenced
by Whitehead Institute. Assembly confirmed by restriction
digest."
67081..67186
/note="Sequence confirmed by overlapping clone AC021003
sequence by Whitehead Institute. Assembly confirmed by
restriction digest."
68186..68322
/note="Sequence confirmed by overlapping clone AC021003
sequence by Whitehead Institute. Assembly confirmed by
restriction digest."
68323..68428
/note="Sequence from overlapping clone AC021003 sequenced
by Whitehead Institute. Assembly confirmed by restriction
digest."
68429..68534
/note="Sequence confirmed by overlapping clone AC021003
sequence by Whitehead Institute. Assembly confirmed by
restriction digest."
76600..76875
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
92981..93194
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
96774..96941
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
99380..99401
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
124439..124492
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
128739..128745
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
130242..130333
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
153887..153985
/note="Sequence from overlapping clone BA347N5
(AL356266). Assembly confirmed by restriction digest."
30906 c 29579 g 47018 t

ORIGIN

Query Match 66.9%; Score 21.4; DB 9; Length 159981;
 Best Local Similarity 80.6%; Pred. No. 3.4e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AAGGTTGGAATTAAGGAATTAATTCGTGA 32
 |||||
 Db 148207 AAGGTTCTTAATGAGAGTAATTAATGTA 148237

RESULT 9
 AC025725
 LOCUS
 DEFINITION
 AC025725 188215 bp DNA linear HTG 12-MAR-2000
 Caenorhabditis elegans clone Y69A2R, *** SEQUENCING IN PROGRESS
 ***, 1 unordered piece.

ACCESSION
 AC025725
 VERSION
 AC025725.1 GI:7230349
 HTG; HTGS PHASE1.
 KEYWORDS
 Caenorhabditis elegans.
 SOURCE
 Caenorhabditis elegans.
 ORGANISM
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 188215)
 Waterston, R.H.
 The sequence of Caenorhabditis elegans clone
 Unpublished
 2 (bases 1 to 188215)
 Waterston, R.H.
 Direct Submission
 Submitted (12-MAR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source
 1. 188215
 /db_xref="taxon:6239"
 /clone="Y69A2R"
 /organism="Caenorhabditis elegans"

BASE COUNT 60803 a 33182 c 33097 g 61133 t

ORIGIN

Query Match 66.9%; Score 21.4; DB 2; Length 188215;
 Best Local Similarity 80.6%; Pred. No. 3.3e+02;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 AAGGTTGGAATTAAGGAATTAATTCGTGA 32
 |||||
 Db 10198 AAGCTTTGAAAAAAGAAATTAATTAATGTA 10228

RESULT 10
 AC084158
 LOCUS
 DEFINITION
 AC084158 188616 bp DNA linear INV 19-APR-2002
 Caenorhabditis elegans cosmid Y69A2AR, complete sequence.
 ACCESSION
 AC084158 GI:10800383
 HTG.
 KEYWORDS
 Caenorhabditis elegans.
 SOURCE
 Caenorhabditis elegans.
 ORGANISM
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 188616)
 Waterston, R.
 Genome sequence of the nematode C. elegans: a platform for

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 99069613
 9851916
 2 (bases 1 to 188616)
 Latreille, P.
 The sequence of C. elegans cosmid Y69A2AR
 Unpublished (2001)
 3 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (13-OCT-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 4 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (28-UN-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 5 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (07-SEP-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (03-NOV-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 188616)
 Waterston, R.
 Direct Submission
 Submitted (19-APR-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 Submitted by:

COMMENT
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: tw@nematode.wustl.edu and jesse@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
www.wormbase.org/db/seq/sequence?name=Y69A2AR;class=Sequence

NEIGHBORING COSMID INFORMATION

The 5' cosmid is Y69A2AL, 200 bp overlap; the 3' cosmid is Y94H6A,
 200 bp overlap.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 Genefinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yui Kohara
http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html and The C.

FEATURES

source

elegans ORFeome cloning project (<http://wormfdb.dfci.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. cRNAs are predicted using the program CRNAScan-SE (Lowe, T.M. and Bddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

Location/Qualifiers

1. 188616

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="Y69A2AR"

complement(2145..2455)

/gene="Y69A2AR.20"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.20); class=S-sequence"

complement(join(2145..2345,2387..2455))

/gene="Y69A2AR.20"

/note="coded for by the following C. elegans CDNAs:

B117493, B117493"

/codon_start=1

/product="Hypothetical protein Y69A2AR.20"

/protein_id="FAL00873.1"

/db_xref="GI:15487619"

/translation="MLALQIQSPSTIWKYIKYQYQVALRISGRPRVTTHMDR

NLRARDEPHRATDIOITISSPNEPVPKRTVRRRLQAGLHG"

4935..10148

/gene="Y69A2AR.19"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.19); class=S-sequence"

join(4935..4974,5029..5094,5313..5459,5695..5841,

6074,6166,6215..6412,6659..8525,8578..8838,8891..9029,

9241..9476,9528..9852,9897..10148)

/gene="Y69A2AR.19"

/note="coded for by the following C. elegans CDNAs:

YK2495.3, YK2495.5, YK33d4.3, YK33d4.5, YK159f10.3,

YK159f10.5, YK251h10.5, YK352f7.5, YK356e6.3, YK356e6.5,

YK420b7.5, YK834g06.3"

/codon_start=1

/product="Hypothetical protein Y69A2AR.19"

/protein_id="AAK68563.2"

/db_xref="GI:16604217"

/translation="MQRKMRFGQLLLALFGVANGYREQNTISSVROKINDNHL

SPRSDEHLMENLDPDASHDFDPLDPELQAGVARQDNNDHLDITRSAGHEL

VBSPLPRDSHDLSTPOELQDVGVARQDQVDPDPSVSLELDIPDTSNNGR

VARQVAPSKSADYLNHTTLIAHTANGIALQAGMKSGIPDAVSELANGSVVSDI

TAPKPDITRLIDKSDTWSHSAIRPKDLPGLTETGSKYKEIIMNDVSPDKLNGP

TAVQPLERLHLVLRYSYIVPVSNTPDITSLNKKVVEQVALAKAGKSMVTSALA

DVGSKPKREHSGFQSGSELKQVLDVDPWPSLANTSAQNNLADGLSFKT

TQGLSHMNDLKPISVSVQOTIAVQVNEISLSBSQENSVAELDKCSRTLP

MDORVAVGELAKHINSIKLSLQPAEFKSTILNSLEKITSFQNSYKSDIDPLA

TATKVAVESIOLKSGSLERKQYLLILOTPADVNNKTYIIGIGIKARITKIS

TREDPITKEBLKVTCLKAGBQSRKLSQAIOATOSIRKDERKLESGASALVSSFA

KDLATIKSIIPETMKKAGVSTDLNGASQAKSDVSGVSSLOQASILESSAIK

ELKSGPEIDGVIKAVSPEDRAITQWNSKEDMSLESGLKAKGPDKIDVSKA

KTMSDYGTPLANKDLDPVKNLEKIKALDAILKALPKPKRATSPDPTVEAKS

TLEKIALADGFNSNKTQOQAPAFAPAFAPLAKPLVTHSSSTGSDSGSITLII

IIVGSLIALGVAVGIVGVRWEKKKAAQOMHEIYVWKAQVYSIBALITIVVAS

LHVWGTOVSVEKSNAYLPKKNRPPLATPCNPETAVMSDGRIRPIHAMVITS

PDVGNATQKRIATQGPLNTSDPFWTVQVQKVEYVVMCQPEKEEKKEHEFPRT

GGVLDLERVYKLTQVTOEILGDSKRTIQVETSKEFPRTITTHQOYMSPOGJPO

GHAQCEDLMNAKESKPIVVCASAGIRVVSIAIOYISAVLARTLVLYLAQVFL

RDQRCAIOTVEQMWVQVGVYRSLSEKNIIMKHYKQEOEMDGGHRTLYLAQVFL

DGRKDYLTLEADGVRKAKODEKATTTGIDBDLGRKKGABCAVAMEY"

complement(10637..11252)

/gene="Y69A2AR.21"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.21); class=S-sequence"

gene

CDS

quence)"

complement(join(10637..10852,10906..11059,11113..11252))

/gene="Y69A2AR.21"

/note="coded for by the following C. elegans CDNAs:

YK11b3.3, YK171h4.5, YK186h5.5, YK257h6.3, YK257h6.5,

YK273g7.3, YK329a9.3, YK329a9.5, YK400a7.3, YK41a10.3,

YK41a10.5, YK418f2.3, YK418f2.5, YK500d4.3, YK500d4.5,

YK508d6.3, YK508d6.5, YK514e4.3, YK514e4.5"

/codon_start=1

/product="Hypothetical protein Y69A2AR.21"

/protein_id="AAK68566.1"

/db_xref="GI:14574457"

/translation="MTDKNNSDSDDEDIESLITNHTSILEAAKRRRLAMKSKIRGI

EMQEDYDEGSTKSKSREVGRFRNHRKDDVAGTQVMDMDLIVOREITHELDVLA

EKADISVDLAMLAPKKIDMDLRDIESKLOKERRQKAVATIRRLAEKGDLMAT

VNAANQNT."

11834..13677

/gene="Y69A2AR.18"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.18); class=S-sequence"

join(11834..11920,11968..12072,12125..12374,12429..12538,

12588..12785,13486..13677)

/gene="Y69A2AR.18"

/note="coded for by the following C. elegans CDNAs:

YK879g02.3, YK879g02.5"

/codon_start=1

/product="Hypothetical protein Y69A2AR.18b"

/protein_id="FAL00872.1"

/db_xref="GI:15487618"

/translation="MLRQAGAGTLQVSGFANAEQARGFATLKDLSIRLSVNIQKI

TKSMKVAAAKYAKARELKGARVGVGAKTPFDNDPVEGVEKQSKOVAVLITS

DRLGCGVSSIVKEAKNLTNNAGKEIVVAILGDSRAGLQRLVANSILISGENEIG

APSPRADASIAKAILDSCYDPEPTGTLIRNRPRTVYSYTSKQIIPLAIRKAKELS

TDSVDDVLOSTSEYSLAQLYIGKESATSLPSKLTLSVPLPLPLFLSLFR

RNVPFHKFSLEPPKRCRDLFFIFLCVTRCYSDEE"

join(11834..11920,11968..12072,12125..12374,12429..12538,

12588..12834,13402..13502)

/gene="Y69A2AR.18"

/note="coded for by the following C. elegans CDNAs:

YK60f11.3, YK60f11.5, YK81e10.5, YK82a12.3, YK82a12.5,

YK83e1.3, YK83e1.5, YK85f6.3, YK85f6.5, YK108f5.3"

/codon_start=1

/product="Hypothetical protein Y69A2AR.18b"

/protein_id="AAK68566.1"

/db_xref="GI:14574457"

/translation="MTDKNNSDSDDEDIESLITNHTSILEAAKRRRLAMKSKIRGI

EMQEDYDEGSTKSKSREVGRFRNHRKDDVAGTQVMDMDLIVOREITHELDVLA

EKADISVDLAMLAPKKIDMDLRDIESKLOKERRQKAVATIRRLAEKGDLMAT

VNAANQNT."

11834..13677

/gene="Y69A2AR.18"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.18); class=S-sequence"

join(11834..11920,11968..12072,12125..12374,12429..12538,

12588..12785,13486..13677)

/gene="Y69A2AR.18"

/note="coded for by the following C. elegans CDNAs:

YK60f11.3, YK60f11.5, YK81e10.5, YK82a12.3, YK82a12.5,

YK83e1.3, YK83e1.5, YK85f6.3, YK85f6.5, YK108f5.3"

/codon_start=1

/product="Hypothetical protein Y69A2AR.18b"

/protein_id="AAK68566.1"

/db_xref="GI:14574457"

/translation="MTDKNNSDSDDEDIESLITNHTSILEAAKRRRLAMKSKIRGI

EMQEDYDEGSTKSKSREVGRFRNHRKDDVAGTQVMDMDLIVOREITHELDVLA

EKADISVDLAMLAPKKIDMDLRDIESKLOKERRQKAVATIRRLAEKGDLMAT

VNAANQNT."

11834..13677

/gene="Y69A2AR.18"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.18); class=S-sequence"

join(11834..11920,11968..12072,12125..12374,12429..12538,

12588..12785,13486..13677)

/gene="Y69A2AR.18"

/note="coded for by the following C. elegans CDNAs:

YK60f11.3, YK60f11.5, YK81e10.5, YK82a12.3, YK82a12.5,

YK83e1.3, YK83e1.5, YK85f6.3, YK85f6.5, YK108f5.3"

/codon_start=1

/product="Hypothetical protein Y69A2AR.18b"

/protein_id="AAK68566.1"

/db_xref="GI:14574457"

/translation="MTDKNNSDSDDEDIESLITNHTSILEAAKRRRLAMKSKIRGI

EMQEDYDEGSTKSKSREVGRFRNHRKDDVAGTQVMDMDLIVOREITHELDVLA

EKADISVDLAMLAPKKIDMDLRDIESKLOKERRQKAVATIRRLAEKGDLMAT

VNAANQNT."

11834..13677

/gene="Y69A2AR.18"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.18); class=S-sequence"

join(11834..11920,11968..12072,12125..12374,12429..12538,

12588..12785,13486..13677)

/gene="Y69A2AR.18"

/note="coded for by the following C. elegans CDNAs:

YK60f11.3, YK60f11.5, YK81e10.5, YK82a12.3, YK82a12.5,

YK83e1.3, YK83e1.5, YK85f6.3, YK85f6.5, YK108f5.3"

/codon_start=1

/product="Hypothetical protein Y69A2AR.18b"

/protein_id="AAK68566.1"

/db_xref="GI:14574457"

/translation="MTDKNNSDSDDEDIESLITNHTSILEAAKRRRLAMKSKIRGI

EMQEDYDEGSTKSKSREVGRFRNHRKDDVAGTQVMDMDLIVOREITHELDVLA

EKADISVDLAMLAPKKIDMDLRDIESKLOKERRQKAVATIRRLAEKGDLMAT

VNAANQNT."

11834..13677

/gene="Y69A2AR.18"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.18); class=S-sequence"

join(11834..11920,11968..12072,12125..12374,12429..12538,

12588..12785,13486..13677)

/gene="Y69A2AR.18"

/note="coded for by the following C. elegans CDNAs:

YK60f11.3, YK60f11.5, YK81e10.5, YK82a12.3, YK82a12.5,

YK83e1.3, YK83e1.5, YK85f6.3, YK85f6.5, YK108f5.3"

/codon_start=1

/product="Hypothetical protein Y69A2AR.18b"

/protein_id="AAK68566.1"

/db_xref="GI:14574457"

/translation="MTDKNNSDSDDEDIESLITNHTSILEAAKRRRLAMKSKIRGI

EMQEDYDEGSTKSKSREVGRFRNHRKDDVAGTQVMDMDLIVOREITHELDVLA

EKADISVDLAMLAPKKIDMDLRDIESKLOKERRQKAVATIRRLAEKGDLMAT

VNAANQNT."

11834..13677

/gene="Y69A2AR.18"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.18); class=S-sequence"

join(11834..11920,11968..12072,12125..12374,12429..12538,

12588..12785,13486..13677)

/gene="Y69A2AR.18"

/note="coded for by the following C. elegans CDNAs:

YK60f11.3, YK60f11.5, YK81e10.5, YK82a12.3, YK82a12.5,

YK83e1.3, YK83e1.5, YK85f6.3, YK85f6.5, YK108f5.3"

/codon_start=1

/product="Hypothetical protein Y69A2AR.18b"

/protein_id="AAK68566.1"

/db_xref="GI:14574457"

/translation="MTDKNNSDSDDEDIESLITNHTSILEAAKRRRLAMKSKIRGI

EMQEDYDEGSTKSKSREVGRFRNHRKDDVAGTQVMDMDLIVOREITHELDVLA

EKADISVDLAMLAPKKIDMDLRDIESKLOKERRQKAVATIRRLAEKGDLMAT

VNAANQNT."

11834..13677

/gene="Y69A2AR.18"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.18); class=S-sequence"

join(11834..11920,11968..12072,12125..12374,12429..12538,

12588..12785,13486..13677)

/gene="Y69A2AR.18"

/note="coded for by the following C. elegans CDNAs:

YK60f11.3, YK60f11.5, YK81e10.5, YK82a12.3, YK82a12.5,

YK83e1.3, YK83e1.5, YK85f6.3, YK85f6.5, YK108f5.3"

/codon_start=1

/product="Hypothetical protein Y69A2AR.18b"

/protein_id="AAK68566.1"

/db_xref="GI:14574457"

/translation="MTDKNNSDSDDEDIESLITNHTSILEAAKRRRLAMKSKIRGI

EMQEDYDEGSTKSKSREVGRFRNHRKDDVAGTQVMDMDLIVOREITHELDVLA

EKADISVDLAMLAPKKIDMDLRDIESKLOKERRQKAVATIRRLAEKGDLMAT

VNAANQNT."

11834..13677

/gene="Y69A2AR.18"

/note="for a graphical representation of this gene see: (www.wormbase.org/db/seq/sequence?name=Y69A2AR.18); class=S-sequence"

join(11834..11920,11968..12072,12125..12374,12429..12538,

12588..12785,13486..13677)

/gene="Y69A2AR.18"

/note="coded for by the following C. elegans CDNAs:

YK60f11.3, YK60f11.5, YK81e10.5, YK82a12.3, YK82a12.5,

YK83e1.3, YK83e1.5, YK85f6.3, YK85f6.5, YK108f5.3"

/codon_start=1

/product="Hypothetical protein Y69A2AR.18b"

/protein_id="AAK68566.1"

/db_xref="GI:14574457"

/translation="MTDKNNSDSDDEDIESLITNHTSILEAAKRRRLAMKSKIRGI

EMQEDYDEGSTKSKSREVGRFRNHRKDDVAGTQVMDMDLIVOREITHELDVLA

EKADISVDLAMLAPKKIDMDLRDIESKLOKERRQKAVATIRRLAEKGDLMAT

VNAANQNT."

JOURNAL Submitted (23-AUG-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
3 (bases 1 to 206320)
REFERENCE Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
AUTHORS Sengulimachak, C., Phelps, K.A., Buckley, D., Raymond, C. and
Haugen, E.D.
TITLE Direct Submission
Submitted (30-APR-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
COMMENT On Apr 30, 2002 this sequence version replaced gi:15281267.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UMG
Web site: <http://www.genome.washington.edu>
Contact: uwgchgs@u.washington.edu
Drafting Center: BCM

----- Project Information

Center project name: chr-3
Center clone name: RP11-1035L22 (bc0587)

----- Summary Statistics

Sequencing vector: M13; 108821; 56% of reads
Chemistry: Dye-terminator; 44% of reads
Chemistry: Dye-terminator; 39% of reads
Chemistry: Dye-terminator; 7% of reads
Chemistry: Dye-terminator; 54% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 206172 bases at least Q40
Consensus quality: 206329 bases at least Q30
Consensus quality: 206318 bases at least Q20
Insert size: 206320; sum-of-contigs
Quality coverage: 6.7% in Q20 bases; sum-of-contigs

----- Overlapping Sequences:

5' Mapping in progress
3' RP11-70P20 (UMGC:bc0189) NC016944

----- Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

----- Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HindIII

BglII

NsiI

SeqDerMap FngPrnt SeqDerMap FngPrnt SeqDerMap FngPrnt

4134 4199 5750 5746 11730 10912
5060

6382	6536	2067	2093	471	<800
512	<800	23113	23465	2365	2357
449	<800	2296	2352	1918	1892
2576	2435	2143	2093	3320	3336
763	<800	213	<800	3896	3912
14581	14639	1292	1270	805	801
511	<800	3330	3324	12	<800
1164	1137	5797	5746	3017	3120
233	<800	73	<800	2190	2175
5383	5502	4684	4673	1134	1110
2447	2435	3485	3483	7391	7328
5965	5912	6554	6644	2204	2175
3370	3384	6939	6843	698	<800
6272	6204	4859	4848	1913	1892
3968	3976	3635	3600	779	<800
2608	2583	6367	6354	2159	2175
2437	2435	113	<800	4727	4679
941	924	4916	4957	11125	10912
1001	997	2479	2581	948	953
4280	4139	16772	16122	522	<800
5303	5502	2964	2988	952	953
355	<800	6115	6171	4238	4193
4289	4497	2628	2767	11374	10912
2260	2251	4465	4493	863	888
2962	2900	395	<800	73	<800
3370	3384	2333	2352	5323	5658
7479	7313	5879	5914	405	<800
460	<800	389	<800	25	<800
1786	1820	4720	4673	1080	1110
9533	9486	372	<800	4217	4088
2233	2251	4480	4493	1913	1892
4983	5101	1783	1922	6083	5979
2490	2435	478	<800	3677	3912
263	<800	1532	1511	54	<800
1324	1369	7151	7155	906	888
218	<800	1634	1701	3963	3912
5060	5101	378	<800	872	888

```

-----
3357 3384 4034 4034 1976 1892
7336 7313 1972 2093 256 <800
767 769 1729 1701 3436 3661
47 <800 1437 1511 1572 1541
2880 2900 685 <800 2872 2903
747 769 1545 1568 6731 6650
6644 6536 2758 2767 4845 4679
2866 2900 2779 2767 308 <800
1837 1820 1781 1701 3311 3326
5645 5502 313 <800 13161 12910
3776 3775 5859 5914 1924 1892
6322 6204 4313 4302 1681 1641
3825 3775 152 <800 5303 5230
1547 1651 1769 1701 881 888
1444 1369 1338 1405 11498 10912
2004 1967 1191 1152 2197 2175
1462 1492 2135 2093 23 <800
-----
Query Match 66.9%; Score 21.4; DB 9; Length 206320;
Best Local Similarity 80.6%; Pred. No. 3.3e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 2 AAGGTTGGAATTAAGGAATTAATTCGTCA 32
Db 153840 AAGGATTGAATTAAGGAATTAATTCGTCA 153870

RESULT 12
AC006892/c 299081 bp DNA linear HTG 26-FEB-1999
LOCUS Caenorhabditis elegans clone Y69A2, *** SEQUENCING IN PROGRESS ***
DEFINITION 10 unordered pieces.
ACCESSION AC006892
VERSION AC006892.2 GI:4309907
KEYWORDS HTG; HTGS PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Bzikavota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
TITLES 1 (bases 1 to 299081)
JOURNAL Waterston, R.H.
REFERENCE The sequence of Caenorhabditis elegans clone
AUTHORS Unpublished
TITLES 2 (bases 1 to 299081)
JOURNAL Waterston, R.H.
REFERENCE Direct Submission
AUTHORS Submitted (24-FEB-1999) Genome Sequencing Center, Washington
TITLES University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL MO 63108, USA
COMMENT On Mar 1, 1999 this sequence; version replaced gi:4263456.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3087: contig of 3087 bp in length
3088 3101: gap of unknown length
3102 7168: contig of 4067 bp in length
7169 7182: gap of unknown length
7183 16968: contig of 9786 bp in length
16969 16982: gap of unknown length
16983 30313: contig of 13311 bp in length
30314 30327: gap of unknown length
30328 45680: contig of 15353 bp in length
45681 45694: gap of unknown length
45695 77792: contig of 32098 bp in length
77793 77806: gap of unknown length
77807 122072: contig of 44266 bp in length
122073 122086: gap of unknown length
122087 157435: contig of 35349 bp in length
157436 157449: gap of unknown length
157450 216837: contig of 59388 bp in length
216838 216851: gap of unknown length
216852 299081: contig of 82230 bp in length.
Location/Qualifiers
1. 299081
/organism="Caenorhabditis elegans"
/db xref="taxon:6239"
/clone="Y69A2"
BASE COUNT 96473 a 52922 c 53657 g 95903 t 126 others
ORIGIN

```

```

Query Match 66.9%; Score 21.4; DB 2; Length 299081;
Best Local Similarity 80.6%; Pred. No. 3.1e+02;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 2 AAGGTTGGAATTAAGGAATTAATTCGTCA 32
Db 35978 AAGGTTGGAATTAAGGAATTAATTCGTCA 35948

```

```

RESULT 13
AR016866/c 1567 bp DNA linear PAT 05-DEC-1998
LOCUS Sequence 100 from patent US 5777200.
DEFINITION AR016866
ACCESSION AR016866
VERSION AR016866.1 GI:3973143
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1567)
AUTHORS Ryals, J.A., Alexander, D.C., Goodman, R.M. and Stinson, J.R.
TITLES Chemically regulatable and anti-pathogenic DNA sequences and uses
 thereof
JOURNAL Patent: US 5777200-A 100 07-JUL-1998;
FEATURES Location/Qualifiers
source 1. 1567
BASE COUNT 469 a 313 c 298 g 477 t 10 others
ORIGIN
Query Match 66.2%; Score 21.2; DB 6; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.1e+02;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 5 GGTTCGAATTAAGGAATTAATTCGTCA 32
Db 454 GGTTCGAATTAAGGAATTAATTCGTCA 427
RESULT 14
AR020892/c 1567 bp DNA linear PAT 05-DEC-1998
LOCUS Sequence 100 from patent US 5789214.
DEFINITION

```

ACCESSION AR020892
VERSION AR020892.1 GI:3975507

KEYWORDS
SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1567)
Unclassified.

AUTHORS Ryals,J.A., Friedrich,L.B., Uknes,S.J. and Ward,E.R.

TITLE Method of inducing gene transcription in a plant

JOURNAL Patent: US 5789214-A 100 04-AUG-1998;

FEATURES
Location/Qualifiers

1..1567

BASE COUNT 469 a 313 c 298 g 477 t 10 others

ORIGIN

Query Match 66.2%; Score 21.2; DB 6; Length 1567;

Best Local Similarity 82.1%; Pred. No. 7.1e+02;

Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 5 GGTTCGAATATAAGGATTAATTCGTGA 32

Db 454 GGTTCGAATATAAGGATTAATTCGTGA 427

RESULT 15

AR027215/c

LOCUS AR027215 1567 bp DNA linear PAT 29-SEP-1999

DEFINITION Sequence 100 from patent US 5856154.

ACCESSION AR027215

VERSION AR027215.1 GI:5938055

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1567)

AUTHORS Ryals,J.A., Alexander,D.C., Goodman,R.M. and Ward,E.R.

TITLE Method of protecting plants from oomycete pathogens

JOURNAL Patent: US 5856154-A 100 05-JAN-1999;

FEATURES
Location/Qualifiers

1..1567

BASE COUNT 469 a 313 c 298 g 477 t 10 others

ORIGIN

Query Match 66.2%; Score 21.2; DB 6; Length 1567;

Best Local Similarity 82.1%; Pred. No. 7.1e+02;

Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 5 GGTTCGAATATAAGGATTAATTCGTGA 32

Db 454 GGTTCGAATATAAGGATTAATTCGTGA 427

Search completed: May 11, 2003, 06:19:46
Job time : 527.927 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 62.2156 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-18
Perfect score: 32
Sequence: 1 caagggtcgaataaagaataattcgtga 32

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	32	21	AAA96852
2	21.2	66.2	1567	20	AAV62796
3	21.2	66.2	1567	20	AAV61680
4	21	65.6	2898	21	AAZ6404
5	20.8	65.0	3103	22	AAH54098
6	20.8	65.0	3438	22	AAH54877
7	20.8	65.0	3463	22	AAH54633
8	20.2	63.1	2894	23	ABL23374
9	20	62.5	360	22	AAF23810

10	20	62.5	1563	22	AAF23809	Escherichia coli s
11	19.8	61.9	49	22	AAV17286	Human ribosomal pr
12	19.8	61.9	241	23	ABV60512	Human prostate exp
13	19.8	61.9	468	22	ABA46292	Human breast cell
14	19.8	61.9	468	22	ABA56837	Human foetal liver
15	19.8	61.9	468	22	ABA26466	Human bone marrow
16	19.8	61.9	468	22	AAK04949	Human brain expres
17	19.8	61.9	468	22	AAK30480	Probe #5025 for ge
18	19.8	61.9	468	22	AAI15092	Probe #5117 used t
19	19.8	61.9	468	22	AAI36431	Probe #4842 used t
20	19.8	61.9	468	22	AAI04651	Human genome-deriv
21	19.8	61.9	468	24	ABV05112	Human cDNA clone (
22	19.8	61.9	4052	23	ABV09300	Drosophila melanog
23	19.8	61.9	466	22	AAK06757	Human brain expres
24	19.2	60.0	466	22	AAK32468	Human bone marrow
25	19.2	60.0	466	22	AAI38302	Probe #6988 used t
26	19.2	60.0	466	22	ABV07254	Human genome-deriv
27	19.2	60.0	3166	24	ABV07038	Listeria monocytog
28	19.2	60.0	11866	5	AAH40141	Sequence of human
29	19.2	60.0	11873	5	AAH40176	Human factor IX (h
30	19.2	60.0	38059	22	AAV54018	Gene #2125 used to
31	19.2	60.0	38059	24	ABV95627	Kidney cancer rela
32	19.2	60.0	267156	24	ABV68560	Listeria monocytog
33	19.2	60.0	2944528	24	ABA03041	DNA encoding novel
34	19.2	60.0	1249	23	AAV27228	Tumour suppressor
35	19	59.4	6277	22	AAV46729	Human immune syste
36	19	59.4	6277	24	ABJ34041	Human immune syste
37	19	59.4	6277	24	ABJ33123	Rat pyruvate kinas
38	19	59.4	11768	24	ABA99851	Human immune syste
39	19	59.4	18434	24	ABJ34006	Drosophila melanog
40	19	59.4	36635	23	ABV07970	Drosophila melanog
41	19	59.4	40388	23	ABJ26342	BAC containing rep
42	19	59.4	94618	21	AAV22285	Pinus radiata cell
43	18.8	58.8	435	21	AAV9604	Oligonucleotide fo
44	18.8	58.8	763	24	ABQ38656	

ALIGNMENTS

RESULT 1
AAA96852
ID AAA96852 standard; DNA; 32 BP.

AC AAA96852;
XX
DT 19-FEB-2001 (first entry)
XX
DE Guide desoxynucleotide building block G4.
XX
KW Promoter, intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; ss.
XX
OS Synthetic.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises

PT		sequence from promoter comprising vascular expression region replaced
PT	-	with sequence from promoter comprising green tissue expression region
XX		
PS	Disclosure; Page 24, 91pp; English.	
XX		
CC	The present sequence represents a guide desoxynucleotide building	
CC	block, which was used to link directional desoxynucleotide building	
CC	blocks during construction of chimeric promoters of the invention.	
CC	The specification describes chimeric expression promoters. These	
CC	chimeric promoters comprise a nucleic acid sequence which is derived	
CC	from a first plant promoter, in which a plant vascular expression	
CC	promoter region is replaced with a nucleic acid sequence derived from	
CC	a second plant promoter comprising a plant green tissue expression	
CC	promoter region. Preferably, the first plant promoter originates from	
CC	Cornellina yellow mottle virus, and the second plant promoter originates	
CC	from the Casseava vein mosaic virus. Especially, the promoters are	
CC	derived from intergenic regions. The chimeric promoters are useful	
CC	for producing transgenic plants.	
XX		
SQ	Sequence 32 BP, 14 A, 3 C, 8 G, 7 T, 0 other;	
Query Match	100.0%; Score 32; DB 21; Length 32;	
Best Local Similarity	100.0%; Pred. No. 0.0023;	
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 CAAGGCTCGAATTAAGGAATAAATTCGTGA 32	
Dn	1 CAAGGCTCGAATTAAGGAATAAATTCGTGA 32	
RESULT: 2		
AAV62796/C		
ID	AAV62796 standard; cDNA; 1567 BP.	
XX		
AC	AAV62796;	
XX		
DT	05-MAR-1999 (first entry)	
XX		
DE	Tobacco SNR CHX independent gene clone 11.3.8.	
XX		
KW	Chemically regulatable DNA promoter; expression control; pesticide;	
XX	herbicide tolerance; systemic acquired resistance gene; ss.	
OS	Nicotiana glauca.	
XX		
PN	US5851766-A.	
XX		
PD	22-DEC-1998.	
PF	31-MAY-1995; 95US-0456262.	
XX		
PR	31-MAY-1995; 95US-0456262.	
XX		
PA	(NOVS) NOVARTIS FINANCE CORP.	
XI		
FI	Harms C, Ryals JA;	
XX		
DR	WI; 1999-080396/07.	
XX		
PT	Isolating chemically regulatable DNA sequences in plants - useful	
PT	for chemically controlling expression in transformed plants	
XX		
PS	Example 40A; Column 255-258; 175pp; English.	
XX		
CC	This sequence represents a clone of the tobacco systemic acquired	
CC	resistance (SAR) cyclohexamide (CHX) independent.	
CC	This gene can be isolated using the method of the invention.	
CC	The method is for isolating a chemically regulatable DNA promoter	
CC	fragment from the 5' flanking region of a chemically regulatable gene in	
CC	a plant tissue. The method allows isolation of sequences which will be	
CC	useful for the controlled expression of genes, under the control of a	
CC	non-coding regulatable sequence. This is useful in plants with a	

CC	herbicide or pesticide detoxification mechanism under the control of a
CC	chemical regulator, the regulator being applied before or with the
CC	herbicide or pesticide to give optimal tolerance. The promoter fragment
CC	is useful for controlling sequences which encode traits such as
CC	height, shape, development, male or female sterility, and the ability
CC	of the plant to withstand cold, heat, salt and drought. The chemical
CC	induction of the promoter allows the regulation of production of
CC	compounds, e.g. flavours, fragrances, pigments, natural sweeteners,
CC	industrial feedstocks, antimicrobials and pharmaceuticals, by
CC	biosynthesis or metabolite conversion, whose biosynthesis is controlled
CC	by endogenous or foreign genes. The method allows control over the time
CC	and rate of gene expression either throughout the whole plant, or in
CC	localized tissues, to achieve e.g. fungal or insect resistance by for
CC	instance dusting the leaves with the chemical regulator. Controlling the
CC	developmental processes by the application of a regulating chemical in
CC	e.g. the commercial production of cultivated crops allows processes such
CC	as germination, flower formation and fruit ripening to be synchronised at
CC	a given time.
XX	
SQ	Sequence 1567 BP; 469 A; 313 C; 298 G; 477 T; 10 other;
OY	Query Match 66.2%; Score 21.2; DB 20; Length 1567;
	Best Local Similarity 82.1%; Pred.No. 50;
Maches	23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
DB	454 GGTTCTAAATTAAGGAAATMAATTGTATTA 427
OY	5 GGTTGGAATTAAGGAATTAATTCGTCA 32
ID	AAV81680 standard; DNA; 1567 BP.
XX	AAV81680;
DT	25-FEB-1999 (first entry)
DE	Tobacco protein-synthetase independent gene p11.3.8.
XX	
KW	Regulation; transcription; plant tissue; chimeric construction; PR;
KM	pathogenesis-related protein; anti-pathogenic; transgenic plant;
XX	beta-1,3-glucanase activity; pest resistance; ss.
OS	Nicotiana sp.
FN	US5847258-A.
PD	08-DEC-1998.
Pf	31-MAY-1995; 95US-0457364.
PR	31-MAY-1995; 95US-0457364.
PR	08-MAR-1988; 88US-0165667.
PR	06-FEB-1989; 89US-0305566.
PR	24-MAR-1989; 89US-0329018.
PR	20-JUN-1989; 89US-0368672.
PR	20-OCT-1989; 89US-0425504.
PR	07-SEP-1990; 90US-0580431.
PR	21-DEC-1990; 90US-0632441.
PR	01-APR-1991; 91US-0678378.
PR	27-SEP-1991; 91US-0768122.
PR	06-MAR-1992; 92US-0848506.
PR	06-NOV-1992; 92US-0973197.
PR	06-APR-1993; 93US-0042847.
PR	12-APR-1993; 93US-0045957.
PR	16-JUL-1993; 93US-0093301.
PR	13-JAN-1994; 94US-0181271.
PA	(NOVS) NOVARTIS FINANCE CORP.
XI	Moyer MB, Payne GB, Ryals JA, Ward ER;
XX	

CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AHS5091 to
 CC AHS5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AHS5091 to
 CC AHS5098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX

SO Sequence 3103 BP; 1088 A; 424 C; 535 G; 1056 T; 0 other;

Query Match 65.0%; Score 20.8; DB 22; Length 3103;
 Best Local Similarity 78.1%; Pred. No. 75;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 CAAGGTTGGAATTAAGATTAATTCGTGA 32
 Db 71 CAATGTTGGAATTAATTAAGATTCGTGA 40

RESULT 6
 AHS4877/c
 ID AHS4877 standard; DNA; 3438 BP.

AC AHS4877;

DT 03-SEP-2001 (first entry)

DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4241.

KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.

OS Staphylococcus epidermidis.

XX MO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000MO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAX) GLAXO GROUP LTD.

XX Kimerly MJ;

XX WPI; 2001-316495/33.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 useful for vaccinating against infections, e.g. endocarditis -

XX Claim 8; Page 1961-1962; 2188pp; English.

CC AHS5204 to AHS53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AHS5091 to
 CC AHS5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AHS5091 to
 CC AHS5098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX

SO Sequence 3438 BP; 1130 A; 608 C; 547 G; 1153 T; 0 other;

Query Match 65.0%; Score 20.8; DB 22; Length 3438;
 Best Local Similarity 78.1%; Pred. No. 75;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy 1 CAAGGTTGGAATTAAGATTAATTCGTGA 32
 Db 1603 CAATGTTGGAATTAATTAAGATTCGTGA 1572

RESULT 7
 AHS4633/c
 ID AHS4633 standard; DNA; 3463 BP.

AC AHS4633;

DT 03-SEP-2001 (first entry)

DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3997.

KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.

OS Staphylococcus epidermidis.

XX MO200134809-A2.

XX 17-MAY-2001.

XX 09-NOV-2000; 2000MO-US30782.

XX 09-NOV-1999; 99US-0164258.

XX (GLAX) GLAXO GROUP LTD.

XX Kimerly MJ;

XX WPI; 2001-316495/33.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 useful for vaccinating against infections, e.g. endocarditis -

XX Claim 8; Page 1670-1671; 2188pp; English.

CC AHS5204 to AHS53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AHS5091 to
 CC AHS5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AHS5091 to
 CC AHS5098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.

CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 3463 BP; 1157 A; 549 C; 641 G; 1116 T; 0 other;

Query Match 65.0%; Score 20.8; DB 22; Length 3463;
Best Local Similarity 78.1%; Pred. No. 76;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CAAGGTTGGAATTAAGCAATTAATTGCTGA 32
DB 2710 CAATTGTTGAAATTAATTAATAAAGATTCGTGA 2679

RESULT 8

ABL23374/C
ID ABL23374 standard; DNA; 2894 BP.

XX ABL23374;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 21595.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200174042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-655660/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -

XX Claim 1; SEQ ID NO 21595; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2894 BP; 698 A; 699 C; 752 G; 745 T; 0 other;

Query Match 63.1%; Score 20.2; DB 23; Length 2894;
Best Local Similarity 88.0%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 GTTCGAAATTAAGCAATTAATTGCT 30
|||||

DB 2568 GTTCGAAATTAATTAATTAATGCT 2544

RESULT 9

AAF23810
ID AAF23810 standard; DNA; 360 BP.

XX AAF23810;

XX 22-MAR-2001 (first entry)

XX Escherichia coli serotype-specific marker fragment.

XX Escherichia coli; serotype-specific marker; O157:H7; O157:NM; O55:H7;

XX serotype detection; ds.

XX Escherichia coli.

XX WO20007247-A1.

XX 21-DEC-2000.

XX 14-JUN-2000; 2000WO-CA00716.

XX 15-JUN-1999; 99US-0139260.

XX (UYGU-) UNIV GUELPH.

XX Chen S, Xu R, Li J;

XX WPI; 2001-112229/12.

XX New DNA markers useful for developing probes, primers and nucleic acid
XX based assays for detecting Escherichia coli serotypes O157:H7, O157:NM
XX and O55:H7 -

XX Example 1; Fig 2; 40pp; English.

XX The present sequence is a part of a novel DNA marker specific for
XX Escherichia coli serotypes O157:H7, O157:NM and O55:H7. The DNA marker is
XX useful for developing probes or primers that can be used to detect
XX E. coli serotypes O157:H7, O157:NM or O55:H7, as well as developing
XX nucleic acid based assays to detect E. coli serotypes in a biological
XX sample. The sample can be a clinical (e.g. blood, urine, serum, tears,
XX saliva, faeces or tissues), food, water or environmental sample.

XX Sequence 360 BP; 94 A; 91 C; 78 G; 97 T; 0 other;

Query Match 62.5%; Score 20; DB 22; Length 360;
Best Local Similarity 82.1%; Pred. No. 1.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 5 GGTTGCAATTAAGCAATTAATTGCTGA 32
DB 159 GGTTCGAAATTAAGCAATTAATTGCTGA 186

RESULT 10

AAF23809
ID AAF23809 standard; DNA; 1583 BP.

XX AAF23809;

XX 22-MAR-2001 (first entry)

XX Escherichia coli serotype-specific 1583 bp marker.

XX Escherichia coli; serotype-specific marker; O157:H7; O157:NM; O55:H7;

XX serotype detection; ds.

XX Escherichia coli.

XX WO20007247-A1.

```

XX 21-DEC-2000.
PD
XX 14-JUN-2000; 2000MO-CA00716.
PE
XX 15-JUN-1999; 99US-0139260.
PR
XX (UYGU-) UNIV GUELPH.
PA
XX Chen S, Xu R, Li J;
PI
XX WPI; 2001-112229/12.
DR
XX
XX New DNA markers useful for developing probes, primers and nucleic acid
PT based assays for detecting Escherichia coli serotypes O157:H7, O157:NM
PT and O55:H7 -
XX
XX Claim 1; Fig 1; 40pp; English.
PS
XX The present sequence is a novel DNA marker specific for Escherichia coli
CC serotypes O157:H7, O157:NM and O55:H7. The DNA marker is useful for
CC developing probes or primers that can be used to detect E. coli serotypes
CC O157:H7, O157:NM or O55:H7, as well as developing nucleic acid based
CC assays to detect E. coli serotypes in a biological sample. The sample can
CC be a clinical (e.g. blood, urine, serum, tears, saliva, faeces or
CC tissues), food, water or environmental sample.
XX
SQ Sequence 1583 BP; 391 A; 430 C; 354 G; 408 T; 0 other;

Query Match      62.5%; Score 20; DB 22; Length 1583;
Best Local Similarity 82.1%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 GGTGGAATTAAGGAATTAATTCTGTA 32
   ||||| ||||| ||||| ||||| |||||
DB 159 GGTGTGAATTAAGGAATTAATTCTTGA 186

RESULT 11
ID AAD17286
AC AAD17286;
XX
XX 29-NOV-2001 (first entry)
DT
XX
DE Human ribosomal protein S24 probe #2 for MRPA assay.
XX
KW Human; multiplex ligation-dependent amplification; amplicon; probe;
KW single nucleotide polymorphism; ribosomal protein; S24; ss.
XX
OS Homo sapiens.
XX
XX WO200161033-A2.
XX
XX 23-AUG-2001.
XX
XX 15-FEB-2001; 2001WO-EP01739.
PF
XX 15-FEB-2000; 2000EP-0200506.
PR
XX (SCHQ/) SCHOUTEN J P.
XX
XX Schouten JP;
PI
XX WPI; 2001-550053/61.
DR
XX
XX An improved multiplex ligation-dependent amplification method for
PT detecting specific single stranded target nucleic acids in samples -
XX
XX Example 2; Page 151; 158pp; English.
XX
XX The invention relates to an improved multiplex ligation-dependent
CC

```

```

CC amplification method for detecting specific single stranded target
CC nucleic acids in samples using a plurality of probe sets comprising at
CC least 2 probes. Each probe comprises a target specific region and a
CC non-complementary region comprising a primer binding site. The probes
CC in each set are ligated when hybridised to a target nucleic acid and
CC amplified by a primer set. The method is used for detecting a nucleotide
CC polymorphism, especially a single nucleotide polymorphism; detecting
CC multiple single stranded target nucleic acid sequences (through the
CC detection of multiple amplicons); determining the absolute or relative
CC abundance of multiple single stranded nucleic acids in a sample; and
CC detection of a break point region in rearranged nucleic acids. By using
CC a femtomolar amount of the probes, a large number of different probe
CC sets can be used to simultaneously detect and quantify a corresponding
CC large number of target sequences with high specificity. The present
CC DNA sequence is human ribosomal protein S24 probe which is used in a
CC multiplex ligatable probe amplification (MLPA) assay.
XX
SQ Sequence 49 BP; 19 A; 7 C; 14 G; 9 T; 0 other;

Query Match      61.9%; Score 19.8; DB 22; Length 49;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 AAGGTTGAAATTAAGGAATTAATTCTGTA 32
   ||||| ||||| ||||| ||||| |||||
DB 10 AAGGTTGAAATTAAGGACAGAAATTCGGA 40

RESULT 12
ID ABV60512
AC ABV60512;
XX
XX 13-SEP-2002 (first entry)
DT
XX
DE Human prostate expression marker cDNA 60503.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200160860-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
PF
XX 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
PI
XX WPI; 2001-662795/76.
DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer; stage of prostate cancer -
XX
XX Claim 1; Page 11524; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC

```

(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
 (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
 (e) selecting a composition for inhibiting prostate cancer in a patient;
 (f) assessing the prostate cell carcinogenic potential of a compound;
 (g) determining whether prostate cancer has metastasized in a patient;
 (h) assessing the aggressiveness or indolence of prostate cancer in a patient;
 (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence 241 BP; 87 A; 37 C; 25 G; 92 T; 0 other;

Query Match 61.9%; Score 19:8; DB 23; Length 241;
 Best Local Similarity 91.3%; Pred. No. 1.5e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

10 GAAATTAAGCAATTAATTGCTGA 32
 |||||
 98 GAAATTAAGCAATTAATTCTTGA 120

RESULT 13
 ID ABA46292/c
 ABA46292 standard; DNA; 468 BP.
 ABA46292;
 01-FEB-2002 (first entry)
 Human breast cell single exon nucleic acid probe #4987.
 Human; microarray; single exon probe; gene expression; breast;
 disease; cancer; ss.
 Homo sapiens.
 WO200157271-A2.
 09-AUG-2001.
 30-JAN-2001; 2001WO-US000662.
 04-FEB-2000; 2000US-0180312.
 26-MAY-2000; 2000US-0207456.
 30-JUN-2000; 2000US-0608408.
 03-AUG-2000; 2000US-0632366.
 21-SEP-2000; 2000US-0234687.
 27-SEP-2000; 2000US-0236359.
 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-496933/54.

New spatially-addressable set of single exon nucleic acid probes, useful for measuring gene expression in sample derived from human breast, comprises number of single exon nucleic acid probes -

Claim 1, SEQ ID NO 4987; 327pp + sequence listing; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BR 474 cells. The method involves contacting the probes with a collection of detectably labelled nucleic acids derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery, and for determining predisposition and/or prognosing breast disease. Gene expression analysis is useful for assessing the toxicity of chemical

CC	agents on cells. The microarray of this invention presents a far greater	CC	agents on cells. The microarray of this invention presents a far greater
CC	diversity of probes for measuring gene expression, with far less bias	CC	diversity of probes for measuring gene expression, with far less bias
CC	than expressed sequence tag microarrays. The method is suitable for	CC	than expressed sequence tag microarrays. The method is suitable for
CC	rapid production of functional information from genomic sequence. The	CC	rapid production of functional information from genomic sequence. The
CC	present sequence is a single exon nucleic acid probe of the invention.	CC	present sequence is a single exon nucleic acid probe of the invention.
CC	Note: The sequence data for this patent did not form part of the	CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly	CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at http://wipo.int/pub/published_pct_sequences .	CC	from WIPO at http://wipo.int/pub/published_pct_sequences .
SQ	Sequence 468 BP; 110 A; 126 C; 52 G; 180 T; 0 other;	SQ	Sequence 468 BP; 110 A; 126 C; 52 G; 180 T; 0 other;
Query Match	61.9%; Score 19.8; DB 22; Length 468;	Query Match	61.9%; Score 19.8; DB 22; Length 468;
Best Local Similarity	77.4%; Pred. No. 1.6e+02;	Best Local Similarity	77.4%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps		Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps	
Oy	2 AAGGTCGAAATTAAGCAATTAATTGCTGA 32	Oy	2 AAGGTCGAAATTAAGCAATTAATTGCTGA 32
Db	230 AATGAGAGAAATTAACGAAAAAATGAGTGA 200	Db	230 AATGAGAGAAATTAACGAAAAAATGAGTGA 200
RESULT 14		RESULT 14	
ABAS56837/C		ABAS56837/C	
ID	ABAS56837 standard; DNA; 468 BP.	ID	ABAS56837 standard; DNA; 468 BP.
XX		XX	
AC	ABA56837;	AC	ABA56837;
XX		XX	
DT	01-FEB-2002 (first entry)	DT	01-FEB-2002 (first entry)
XX		XX	
DE	Human foetal liver single exon nucleic acid probe #5142.	DE	Human foetal liver single exon nucleic acid probe #5142.
XX		XX	
KX	Human; foetal liver; gene expression; single exon nucleic acid probe; ss	KX	Human; foetal liver; gene expression; single exon nucleic acid probe; ss
XX		XX	
OS	Homo sapiens.	OS	Homo sapiens.
XX		XX	
PN	WO200157277-A2.	PN	WO200157277-A2.
XX		XX	
PD	09-AUG-2001.	PD	09-AUG-2001.
XX		XX	
PF	30-JAN-2001; 2001WO-US00669.	PF	30-JAN-2001; 2001WO-US00669.
XX		XX	
PR	04-FEB-2000; 2000US-0180312.	PR	04-FEB-2000; 2000US-0180312.
XX	26-MAY-2000; 2000US-0207456.	XX	26-MAY-2000; 2000US-0207456.
FR	30-JUN-2000; 2000US-0608408.	FR	30-JUN-2000; 2000US-0608408.
XX	03-AUG-2000; 2000US-0632366.	XX	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.	PR	21-SEP-2000; 2000US-0234687.
XX	27-SEP-2000; 2000US-0236359.	XX	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.	PR	04-OCT-2000; 2000GB-0024263.
XX		XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.	PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX		XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;	PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX		XX	
DR	WPI; 2001-483447/52.	DR	WPI; 2001-483447/52.
XX		XX	
PT	Human genome-derived single exon nucleic acid probes useful for	PT	Human genome-derived single exon nucleic acid probes useful for
XX	analyzing gene expression in human foetal liver -	XX	analyzing gene expression in human foetal liver -
PS	Claim 1; SEQ ID NO 5142; 639pp + sequence listing; English.	PS	Claim 1; SEQ ID NO 5142; 639pp + sequence listing; English.
CC	The invention relates to a single exon nucleic acid probe for	CC	The invention relates to a single exon nucleic acid probe for
CC	measuring human gene expression in a sample derived from human foetal	CC	measuring human gene expression in a sample derived from human foetal
CC	liver. The single exon nucleic acid probes may be used for predicting,	CC	liver. The single exon nucleic acid probes may be used for predicting,
CC	measuring and displaying gene expression in samples derived from human	CC	measuring and displaying gene expression in samples derived from human
CC	foetal liver. The present sequence is a single exon nucleic acid	CC	foetal liver. The present sequence is a single exon nucleic acid
CC	probe of the invention.	CC	probe of the invention.
CC	Note: The sequence data for this patent did not form part of the	CC	Note: The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly	CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at http://wipo.int/pub/published_pct_sequences .	CC	from WIPO at http://wipo.int/pub/published_pct_sequences .
XX		XX	
SQ	Sequence 468 BP; 110 A; 126 C; 52 G; 180 T; 0 other;	SQ	Sequence 468 BP; 110 A; 126 C; 52 G; 180 T; 0 other;
Query Match	61.9%; Score 19.8; DB 22; Length 468;	Query Match	61.9%; Score 19.8; DB 22; Length 468;
Best Local Similarity	77.4%; Pred. No. 1.6e+02;	Best Local Similarity	77.4%; Pred. No. 1.6e+02;

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 12.8321 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-18

Perfect score: 32
Sequence: 1 caagggtcgaataaagaataatcgtga 32

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	21.2	66.2	1567	1 US-08-181-271A-100	Sequence 100, App
C 2	21.2	66.2	1567	1 US-08-449-315-100	Sequence 100, App
C 3	21.2	66.2	1567	1 US-08-444-803-100	Sequence 100, App
C 4	21.2	66.2	1567	1 US-08-449-043-100	Sequence 100, App
C 5	21.2	66.2	1567	1 US-08-456-265A-100	Sequence 100, App
C 6	21.2	66.2	1567	1 US-08-455-416-100	Sequence 100, App
C 7	21.2	66.2	1567	1 US-08-455-244-100	Sequence 100, App
C 8	21.2	66.2	1567	1 US-08-454-876-100	Sequence 100, App
C 9	21.2	66.2	1567	2 US-08-457-364-100	Sequence 100, App
C 10	21.2	66.2	1567	2 US-08-456-262-100	Sequence 100, App
C 11	21.2	66.2	1567	2 US-08-455-340-100	Sequence 100, App
C 12	21.2	66.2	1567	2 US-08-455-736-100	Sequence 100, App
C 13	21.2	66.2	1567	2 US-08-971-217-100	Sequence 100, App
C 14	21.2	66.2	1567	4 US-09-350-600-100	Sequence 100, App
C 15	20	62.5	31880	4 US-09-453-702B-242	Sequence 242, App
C 16	18.8	58.8	1683	4 US-09-009-443-6	Sequence 6, Appli
C 17	18.8	58.8	4307	3 US-08-944-594-1	Sequence 1, Appli
C 18	18.6	58.1	4550	4 US-09-103-663-35	Sequence 35, Appli
C 19	18.4	57.5	1560	1 US-08-194-180-2	Sequence 2, Appli
C 20	18.2	56.9	5361	4 US-09-276-531-132	Sequence 132, App
C 21	18.2	56.9	5361	4 US-08-973-462-2	Sequence 2, Appli
C 22	18.2	56.9	6152	4 US-08-973-462-1	Sequence 1, Appli
C 23	17.8	55.6	429	4 US-09-641-638-412	Sequence 412, App
C 24	17.8	55.6	429	4 US-08-641-638-413	Sequence 413, App
C 25	17.6	55.0	943	1 US-08-473-157A-3	Sequence 3, Appli
C 26	17.6	55.0	2049	1 US-09-134-001C-2413	Sequence 2413, App
C 27	17.6	55.0	3431	4 US-09-155-078-1	Sequence 1, Appli

C 28	17.6	55.0	21338	4 US-08-961-527-20	Sequence 20, Appli
C 29	17.4	54.4	2135	3 US-08-965-903B-7	Sequence 7, Appli
C 30	17.4	54.4	2509	2 US-08-954-333-6	Sequence 6, Appli
C 31	17.4	54.4	2797	2 US-08-954-333-8	Sequence 8, Appli
C 32	17.4	54.4	4155	1 US-07-876-280-1	Sequence 1, Appli
C 33	17.4	54.4	4155	1 US-07-675-772-1	Sequence 1, Appli
C 34	17.4	54.4	4155	1 US-08-063-170-1	Sequence 1, Appli
C 35	17.4	54.4	4155	1 US-08-158-232-1	Sequence 1, Appli
C 36	17.4	54.4	4155	1 US-08-304-626-1	Sequence 1, Appli
C 37	17.4	54.4	4155	1 US-08-316-301A-1	Sequence 1, Appli
C 38	17.4	54.4	4155	1 US-08-611-928-1	Sequence 1, Appli
C 39	17.4	54.4	4155	3 US-09-173-891-1	Sequence 1, Appli
C 40	17.4	54.4	4155	4 US-09-076-137-1	Sequence 1, Appli
C 41	17.4	54.4	4155	5 PCT-US92-03624-1	Sequence 1, Appli
C 42	17.4	54.4	4155	6 PCT-US92-03624-1	Sequence 1, Appli
C 43	17.4	54.4	4155	6 PCT-US92-03624-1	Sequence 1, Appli
C 44	17.4	54.4	9048	3 US-08-973-273-4	Sequence 4, Appli
C 45	17.2	53.8	1575	1 US-08-375-709-18	Sequence 18, Appli

ALIGNMENTS

RESULT 1
US-08-181-271A-100/c
Sequence 100, Application US/08181271A
Patent No. 5614395
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Weins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Meyer, Mary B.
APPLICANT: Neuhau, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Umes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Sherice C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/181,271A
FILING DATE: 13-JAN-94
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-181-271A-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCTTCGAATTAAGGATTAATTGCTGA 32
DB 454 GGTCTTAATTAAGGATTAATTGATTA 427

RESULT 2
US-08-449-315-100/c
Sequence 100, Application US/08449315
Patent No. 5650505
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.

APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Speirson, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/449,315
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-449-315-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTTCGAATTAAGCAATTAATTCGTGA 32
DB 454 GGTCTAAATTAAGCAATTAATTAATTA 427

RESULT 3
US-08-444-803-100/c
Sequence 100, Application US/08444803
Patent No. 5654414
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Utnes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,803
FILING DATE: 19-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378

FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-444-803-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTTCGAATTAAGCAATTAATTCGTGA 32
DB 454 GGTCTAAATTAAGCAATTAATTAATTA 427

RESULT 4
US-08-449-043-100/c
Sequence 100, Application US/08449043
Patent No. 5689044
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice

APPLICANT: Moyer, Mary B.
APPLICANT: Neuhau, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Spertison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Ukenes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESS: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,043
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,432
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-449-043-100
Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 5 GGTCGAATTAAGATTAATTCGTGA 32
DB 454 GGTTCTAATTAAGATTAATTCATTA 427
RESULT 5
US-08-456-265A-100/c
Sequence 100, Application US/08456265A
Patent No. 5767369
GENERAL INFORMATION:
APPLICANT: Alexander, Danny C.
APPLICANT: Ryals, John A.
APPLICANT: Goodman, Robert M.
APPLICANT: Stinson, Jeffrey R.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESS: CIBA-GEIGY Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,265A
FILING DATE: 31-MAY-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/181,271
FILING DATE: 13-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727/DIV10
TELEPHONE: (919)541-8687
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
US-08-456-265A-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 5 GGTTCGAATTAAGCAATTAATTCGTCA 32
Db 454 GGTTCGAATTAAGCAATTAATTCGTCA 427

RESULT 6
US-08-455-416-100/c
Sequence 100, Application US/08455416
Patent No. 5777200
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyet, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Spertson, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,416
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-08-455-416-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTCGAATTAAGGAATTAATTCGTGA 32
|||||
DB 454 GGTTCTAATTAAGGAATTAATTCGTTA 427

RESULT 7

US-08-455-244-100/c
Sequence 100, Application US/08455244
Patent No. 5789214

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedlich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Melns, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaue, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,244
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-455-244-100

Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTCGAATTAAGGAATTAATTCGTGA 32
|||||
DB 454 GGTTCTAATTAAGGAATTAATTCGTTA 427

RESULT 8

US-08-454-876-100/c
Sequence 100, Application US/08454876
Patent No. 5804693

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedlich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Melns, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaue, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/454,876
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PI/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-454-876-100
Query Match 66.2%; Score 21.2; DB 1; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 5 GGTTCGAATTAAGATTAATTCGTA 32
DB 454 GGTTCGAATTAAGATTAATTCGTA 427
RESULT 9
US-08-457-364-100/c
Sequence 100, Application US/08457364
Patent No. 5847258
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhau, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Unnes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,364
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 800
FILING DATE: 08/181,271
APPLICATION NUMBER: US 08/093,301
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 07/937,197
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847

;; FILING DATE: 6-APR-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/632,441
;; FILING DATE: 21-DEC-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/425,504
;; FILING DATE: 20-OCT-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/848,506
;; FILING DATE: 6-MAR-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/768,122
;; FILING DATE: 27-SEP-1991
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/580,431
;; FILING DATE: 7-SEP-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/368,672
;; FILING DATE: 20-JUN-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/329,018
;; FILING DATE: 24-MAR-1989
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/045,957
;; FILING DATE: 12-APR-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Elmer, James Scott
;; REGISTRATION NUMBER: 36,129
;; REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919)541-8614
;; TELEFAX: (919)541-8619
;; INFORMATION FOR SEQ ID NO: 100:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1567 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-457-364-100
Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Cy 5 GGTTCGATTAAGGATTAATTCGTGA 32
Db 454 GGTTCGATTAAGGATTAATTCGTGA 427
RESULT 10
US-08-456-262-100/C
Sequence 100, Application US/08456262
Patent No. 5851766
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedlich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Helms, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhau, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC

;; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
;; NUMBER OF SEQUENCES: 106
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CIBA-GEIGY Corporation
;; STREET: 7 Skyline Drive
;; CITY: Hawthorne
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,262
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8619
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-456-262-100

Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTGGAATTAAGGATTAATTCGTGA 32
DB 454 GGTCTAATAATTAAGGATTAATTCATTA 427

RESULT 11
US-08-456-240-100/c
Sequence 100, Application US/08456240
Patent No. 5856154
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,240
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
PRIOR APPLICATION DATA:
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-456-240-100

Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GGTGGAATTAAGGATTAATTCGTGA 32
DB 454 GGTCTAATAATTAAGGATTAATTCATTA 427

RESULT 12
US-08-455-736-100/c
Sequence 100, Application US/08455736
Patent No. 5880328
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.

TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,736
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-1994
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-455-736-100
Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. NO. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 5 GGTTCGAATTAAGGATTAATTCGTCA 32
Db 454 GGTTCGAATTAAGGATTAATTCGTCA 427
RESULT 13
US-08-971-217-100/C
Sequence 100, Application US/08971217
Patent No. 5942662
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Harms, Christian
APPLICANT: Friedrich, Leslie
APPLICANT: Beck, James
APPLICANT: Uknes, Scott
TITLE OF INVENTION: INDUCIBLE HERBICIDE RESISTANCE
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5942662artis Corporation
STREET: 3054 Cornwallis Road, P.O. Box 12257
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,217
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,364
FILING DATE: 31-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/181,271
FILING DATE: 13-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504

FILED DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727/DIV5/CONF
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-971-217-100
MOLECULE TYPE: DNA (genomic)
Query Match 66.2%; Score 21.2; DB 2; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCTCGAATAAGGATTAATTCGTGA 32
DB 454 GGTCTTAATAAGGATTAATTCGTGA 427

RESULT 14
US-09-350-600-100/c
Sequence 100, Application US/09350600
GENERAL INFORMATION:
APPLICANT: Meigs, Frederick
APPLICANT: Shinshi, Hideaki
APPLICANT: Wenzler, Herman
APPLICANT: Hofsteenge, Jan
APPLICANT: Ryals, John
APPLICANT: Sperisen, Christoph
TITLE OF INVENTION: DNA SEQUENCES ENCODING POLYPEPTIDES
TITLE OF INVENTION: HAVING BETA-1,3-GLUCANASE ACTIVITY
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 626234artis Corporation
STREET: 3054 Cornwallis Road, P.O. Box 12257
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/350,600
FILING DATE:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/971,217
FILING DATE: 14-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/457,364
FILING DATE: 31-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/181,271
FILING DATE: 13-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/381,443
FILING DATE: 18-JUL-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/353,312
FILING DATE: 17-MAY-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/226,303
FILING DATE: 29-JUL-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-198250
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 1567 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
US-09-350-600-100

Query Match 66.2%; Score 21.2; DB 4; Length 1567;
Best Local Similarity 82.1%; Pred. No. 7.4;
Matches 23; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 5 GGTCGAATTAAGGAATTAATTCGTGA 32
|||||
Db 454 GGTTCTAATTAAGGAATTAATTCATTA 427

RESULT 15

US-09-453-702B-242/c
Sequence 242. Application US/09453702B
Patent No. 6365723

GENERAL INFORMATION:

APPLICANT: Blactner, Frederick R.
Perna, Nicole T.
Burland, Valerie
Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pluckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 242:

SEQUENCE CHARACTERISTICS:

LENGTH: 31880

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 242:

US-09-453-702B-242

Query Match 62.5%; Score 20; DB 4; Length 31880;

Best Local Similarity 82.1%; Pred. No. 30;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 5 GGTCGAATTAAGGAATTAATTCGTGA 32
|||||
Db 28206 GGTTGTGAATTAAGGAATTAATTCGTGA 28179

Search completed: May 11, 2003, 03:08:10
Job time : 17.8321 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 514.406 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-18

Perfect score: 32
Sequence: 1 caagggtcgaataaagaataaattcgtga 32

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST.*
1: em_estba:*
2: em_esthum:*
3: em_esthu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gse:*
18: em_gse_hum:*
19: em_gse_inv:*
20: em_gse_pla:*
21: em_gse_vrt:*
22: em_gse_fun:*
23: em_gse_man:*
24: em_gse_mus:*
25: em_gse_other:*
26: em_gse_pro:*
27: em_gse_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	71.9	728	17	AQ449597 500002C12
2	22.4	70.0	144	9	AV174753 AV174753
3	21.6	67.5	1016	17	BH147730 BH147730
4	21.4	66.9	309	10	BB490355 BB490355
5	21.4	66.9	375	17	AZ478586 AZ478586
6	21.4	66.9	432	13	BI327328 BI327328

7	21	65.6	1074	17	CNS06XG0	AL119686 T3 end of
8	20.8	65.0	365	17	AQ100322	AQ100322 HS 3054_A
9	20.8	65.0	502	17	AQ214572	AQ214572 HS 3218_B
10	20.8	65.0	546	13	BI746889	BI746889 tm3c01_Y
11	20.8	65.0	560	17	AQ237213	AQ237213 RPI111-65
12	20.8	65.0	662	17	AQ239775	AQ239775 RPI111-64
13	20.8	65.0	669	17	CNS045XL	AL275970 Tetradon
14	20.8	65.0	692	17	AG144326	AG144326 Pan treg1
15	20.8	65.0	768	10	AV853564	AV853564 AV853564
16	20.8	65.0	900	17	AQ746633	AQ746633 HS 2278_A
17	20.6	64.4	422	17	AQ734580	AQ734580 HS 3012_A
18	20.6	64.4	467	10	BE235639	BE235639 143201_MA
19	20.6	64.4	532	13	BI742345	BI742345 kt3Be09_Y
20	20.6	64.4	573	10	BE235635	BE235635 143196_MA
21	20.6	64.4	1160	17	AZ936007	AZ936007 ODG687_O1
22	20.4	63.7	445	17	AQ664107	AQ664107 HS 5486_A
23	20.4	63.7	449	10	AM864304	AM864304 PM4-SN001
24	20.4	63.7	479	10	BB829456	BB829456 BB829456
25	20.4	63.7	484	10	BB829743	BB829743 BB829743
26	20.4	63.7	535	17	AQ429861	AQ429861 HS 5080_A
27	20.4	63.7	639	12	BE885951	BE885951 601507582
28	20.4	63.7	1066	17	CNS02CEH	AL196226 Tetradon
29	20.2	63.1	218	9	AV078828	AV078828 AV078828
30	20.2	63.1	630	17	AZ198427	AZ198427 SP_1037_B
31	20.2	63.1	697	17	AZ360780	AZ360780 IM0104L08
32	20.2	63.1	847	12	BE866488	BE866488 601678853
33	20.2	63.1	971	17	CNS07DG5	AL440427 T3 end of
34	20.2	63.1	1031	17	CNS00JB7	AL075957 Drosophila
35	20.2	62.5	145	17	BH408228	BH408228 100705610
36	20.2	62.5	296	9	A1059555	A1059555 UI-R-C1-1
37	20.2	62.5	507	12	BF549568	BF549568 UI-R-C1-1
38	20.2	62.5	531	10	BE082568	BE082568 RC5-BT064
39	20.2	62.5	551	17	AQ838015	AQ838015 HS 4708_A
40	20.2	62.5	606	17	BH766421	BH766421 BMEAC3A1E
41	20.2	62.5	607	17	AZ612822	AZ612822 IM0439H23
42	20.2	62.5	776	13	BI656752	BI656752 603281423
43	20.2	62.5	803	9	AL663807	AL663807 AL663807
44	20.2	62.5	867	12	BF675988	BF675988 602083948
45	20.2	62.5	1226	14	BQ677337	BQ677337 AGENCOURT

ALIGNMENTS

RESULT 1
LOCUS AQ449597/c 728 bp DNA linear GSS 08-APR-1999
DEFINITION 500002C12.x2 CpIOWAM13mp18gDNA1 Cryptosporidium parvum genomic, DNA
sequence.
ACCESSION AQ449597
VERSION AQ449597.1 GI:4578734
KEYWORDS GSS.
SOURCE Cryptosporidium parvum.
ORGANISM Cryptosporidium parvum
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Cryptosporidiidae; Cryptosporidium.
REFERENCE 1 (bases 1 to 728)
AUTHORS Hyman,R.W., Fung,E., Qin,F., Rowley,D. and Davis,R.W.
TITLE Cryptosporidium parvum genome sequencing demonstration project
JOURNAL Unpublished (1999)
COMMENT Contact: Hyman, R. W.
Stanford DNA Sequencing and Technology Center
Stanford University School of Medicine, Palo Alto
855 California Avenue, Palo Alto, CA 94304, USA
Tel: 650 812 1972
Fax: 650 812 1975
Email: hyman@sequence.stanford.edu
For Annotation Data see http://medefgh.ucsf.edu/id/CpTags/home.html
Seq primer: M13(-21) forward
Class: shotgun.
FEATURES
Location/Qualifiers
1..728
/organism="Cryptosporidium parvum"

[illegible]

/note="Organ: mammary gland; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTGACCAATCTGAAGTGAGCGCCGCCAATGCTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bernaldo."

BASE COUNT 32 a 39 c 22 g 51 t

ORIGIN

Query Match 70.0%; Score 22.4; DB 9; Length 144;

Best Local Similarity 81.2%; Pred. No. 3e+03;

Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

CY 1 CAAGGGTTCGAATTAAGGATAAATTCGTGA 32
 ||||| |||||
DB 114 CAAGGGGTATTAAATGATGAAACAATTCGTGA 83

RESULT 3

LOCUS BH147730 1016 bp DNA linear GSS 27-AUG-2001

DEFINITION ENT0L82TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomicC, DNA sequence.

ACCESSION BH147730

VERSION BH147730.1 GI:15305567

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 1016)

AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)

JOURNAL COMMENT Unpublished (2001)

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Seq primer: ML3-Reverse

Class: Shotgun

High quality sequence start: 16

High quality sequence stop: 492.

location/Qualifiers

1. 1016

 /organism="Entamoeba histolytica"

 /strain="HMI:IMSS"

 /db_xref="taxon:5759"

 /clone_lib="Entamoeba histolytica Sheared DNA"

 /note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Recombinant DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999). "

BASE COUNT 387 a 147 c 175 g 307 t

Query Match 67.5%; Score 21.6; DB 17; Length 1016;
 Best Local Similarity 85.7%; Pred. No. 2.5e+03;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAGGTTCCGAATTAAGATAATTC 28
 Db 962 CAGGGTTGGAGAAAAAGATAATTC 989

RESULT 4
 BB490355/c 309 bp mRNA linear EST 25-JUL-2000

LOCUS BB490355 RIKEN full-length enriched, 13 days embryo stomach Mus
 DEFINITION musculus cDNA clone D530014M20 3', mRNA sequence.

ACCESSION BB490355
 VERSION BB490355.1 GI:9448982

KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 309)
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, P., Ishii, Y., Ishikawa, D., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, P., Tomihata, N., Toyama, T., Tsunoda, Y., Watanabe, S., Yamane, T., Yamashita, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Yamamatsu, M. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Konno, H., et al.)
 JOURNAL Unpublished (2000)
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermolabile and thermoactivated enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunari, T., Akiyama, U., Shibata, K., Iizawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 13-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
 source Location/Qualifiers

1..309
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="D530014M20"
 /clone_1ib="RIKEN full-length enriched, 13 days embryo stomach"
 /issue_type="stomach"
 /dev_stage="13 days embryo"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia"

BASE COUNT
 ORIGIN 83 a 80 c 57 g 88 t 1 others

Query Match 66.9%; Score 21.4; DB 10; Length 309;
 Best Local Similarity 80.6%; Pred. No. 4.4e+03;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 CAAGGTTCCGAATTAAGATAATTCG 31
 Db 61 CAGGGTTGGAGAAAAAGATAATTCG 31

RESULT 5
 AZ478586/c 375 bp DNA linear GSS 04-OCT-2000

LOCUS AZ478586 RIKEN full-length enriched, 13 days embryo stomach Mus
 DEFINITION mouse 10kb plasmid UUGC1M library Mus musculus genomic
 clone UUGC1M0298G17 R, DNA sequence.

ACCESSION AZ478586
 VERSION AZ478586.1 GI:10637592

KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 375)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu

Insert length: 10000 Std Error: 0.00
 Plate: 0298 row: G column: 17
 Seq primer: CACACAGAAACGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 375.
 Location/Qualifiers

FEATURES
 source

1..375
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC1M0298G17"
 /clone_1ib="Mouse 10kb plasmid UUGC1M library"
 /sex="Male"
 /lab_host="B. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD29v. Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

was blunt-ended and repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g1473211[gb|AF129072.1]), a copy-number inducible derivative of plasmid λ 1, the vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptorised mouse DNA was annealed to adaptorised vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT	129 a	77 c	66 g	103 t
ORIGIN				

Query Match	66.9%	Score 21.4	DB 17	Length 375
Best Local Similarity	80.6%	Pred. No. 4.1e+03		
Matches 25; Conservative	0	Mismatches 6	Indels 0	Gaps 0

```

Oy      1  CAAGGCTTCGAATAAAGGATTAATTCGTG  31
          |||||
Db      356 CAAGTGTTCGAATAAAGGATTAAGGTGTG  326

```

RESULT 6	BI327328/c	LOCUS	DEFINITION
BI327328	432 bp	Porcine Thymus cDNA library Sus scrofa cDNA, mRNA	EST 01-SEP-2001
AR077409THYBH095	sequence.		

VERSION	BI327328.1	GI:15417031
KEYWORDS	EST.	
SOURCE	plg.	

ORGANISM	REFERENCE
<i>Sus scrofa</i>	1 (bases 1 to 432)
<i>Escherichia</i>	1 (bases 1 to 432)
<i>Mammalia</i>	1 (bases 1 to 432)
<i>Rink A.</i>	1 (bases 1 to 432)
<i>Santuschl, E.M. and Baatle, C.W.</i>	1 (bases 1 to 432)
<i>Amplified, Normalized cdna Libraries from a Porcine Model of Orthopedic Implant Associated Staphylococcus aureus Infection</i>	1 (bases 1 to 432)
<i>Orthopedic Implant Associated Staphylococcus aureus Infection</i>	1 (bases 1 to 432)
<i>unpublished (2001)</i>	1 (bases 1 to 432)

COMMENT
Contact: Rink A.
Department of Animal Biotechnology
College of Agriculture, Biotechnology and Natural Resources,
University of Nevada, Reno
MS 202, PA 103, 1664 N Virginia St., Reno, NV 89557-0236, USA
Tel: 775 784 1705
Fax: 775 784 1375

Email: arink@cabmr.unr.edu
Tissues and cells are derived from a porcine model for
implant-associated infection using 1000 cfu of *Staphylococcus*
aureus in a tibial transection, reduced and internally fixed with a
dynamic compression plate. NOTE: The sequences contain a 'cDNA
adapter' between the EcoRI site and the start of the 5ST. The
adapter sequence is 'AATTGGCAGCGAG'.

```
FEATURES
source
Location/Qualifiers
1. 432
/organism="Sus scrofa"
/strain="crossbred"
/db_xref="taxon:9623"
/clone_lib="Porcine Thymus cDNA library"
/cisue_type="Thymus"
```

```

/cell_type="mixed"
/dev_stage="control, 5 month old castrated male"
/lab_host="SOLR"
/notes="Vector: pBSK, Site_1: Eco RI, Site_2: XhoI, Tissues
and cells are derived from a porcine model for
implant-associated infection using 1000 cfu of
Staphylococcus aureus in a tibial transection, reduced and
internally fixed with a dynamic compression plate. NOTE:
The sequences contain a 'CDNA adapter' between the EcoRI
site and the start of the 5ST. The adapter sequence is

```

	'AATTCGGCAGAG', "				
BASE COUNT	148 a	97 c	78 g	100 t	9 others
ORIGIN					

Query Match	66.9%	Score 21.4;	DB 13;	Length 432;
Best Local Similarity	80.6%	Pred. No. 3.9e+03;		
Matches 25; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0;

QY 2 AAGGTTCCGAATAAAGCAATAAATTCGTGA 32
 ||||| ||||| ||||| | |||
Db 393 AAGGATTCCGAATAAAGCAATAATTACGGGA 363

RESULT 7	CNS06XG0	LOCUS	DEFINITION
	CNS06XG0	1074 bp	DNA linear GSS 06-JUL-2001
		T3 end of clone XAX0AA002012	of library XAX0AA from strain CBS 7064
		of <i>Pichia farinosa</i> , genomic survey sequence.	

KEYWORDS	GSS.
SOURCE	Pichia farinosa.
ORGANISM	Pichia farinosa

[illegible]

TITLE
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE	20584711
PUBMED	1152876
REFERENCE	2 (bases 1 to 1074)
AUTHORS	de Montigny, J., Spehner, C., Soucier, J., Tekala, F., Dujon, B., Wincker, P., Artiguenave, F. and Follet, S.
TITLE	Genomic exploration of the hemiascomycetous yeasts: 15. <i>Pichia sorbitophila</i>

JOURNAL	FEBs Letc. 487 (1), 87-90 (2000)
MEDLINE	20584725
PUBMED	11152890
REFERENCE	3 (bases 1 to 1074)
AUTHORS	Genoscope
TITLE	Direct Submission
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequençage,

COMMENT

2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seer@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*

FEATURES
source
1..1074
Location/Qualifiers

Query Match	65.6%	Score 21	DB 17	Length 1074
-------------	-------	----------	-------	-------------

Best Local Similarity 82.8%; Pred. No. 3.7e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 920 CAAGGCTTGAAGATAAGATTAATTTG 948

QY 1 CAAGGCTTGAAGATAAGATAATTCG 29
|||||

RESULT 8
AQ100322 365 bp DNA linear GSS 27-AUG-1998
LOCUS
DEFINITION HS-3054_A1_D06_MF CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3054 Col=11 Row=G, DNA sequence.
ACCESSION AQ100322
VERSION AQ100322.1 GI:3471351
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
AUTHORS Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL MEDLINE
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3054 row: G column: 11
Class: BAC ends
High quality sequence stop: 365.
FEATURES
source
1.365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3054 Col=11 Row=G"
/clone_1b="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"
BASE COUNT 114 a 67 c 74 g 110 t
ORIGIN

Query Match 65.0%; Score 20.8; DB 17; Length 365;
Best Local Similarity 78.1%; Pred. No. 6.3e+03;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 83 CAAGGCTTGAAGATAAGATAATTTCTTA 114
|||||

QY 1 CAAGGCTTGAAGATAAGATAATTCGTA 32
|||||

Db 83 CAAGGCTTGAAGATAAGATAATTTCTTA 114
|||||

RESULT 9
AQ214572 502 bp DNA linear GSS 18-SEP-1998
LOCUS
DEFINITION HS-3318_B2_G02_T7 CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3218 Col=4 Row=N, DNA sequence.
ACCESSION AQ214572
VERSION AQ214572.1 GI:3625773
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 502)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL MEDLINE
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3218 row: N column: 4
Class: BAC ends
High quality sequence stop: 502.
FEATURES
source
1.502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3218 Col=4 Row=N"
/clone_1b="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"
BASE COUNT 147 a 91 c 105 g 158 t 1 others
ORIGIN

Query Match 65.0%; Score 20.8; DB 17; Length 502;
Best Local Similarity 78.1%; Pred. No. 5.6e+03;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Db 209 CAAGGCTTGAAGATAAGATAATTTCTTA 240
|||||

QY 1 CAAGGCTTGAAGATAAGATAATTCGTA 32
|||||

Db 209 CAAGGCTTGAAGATAAGATAATTTCTTA 240
|||||

RESULT 10
BT746889 546 bp mRNA linear EST 25-SEP-2001
LOCUS
DEFINITION tm33c01.y1 Meloidogyme arenaria egg pMMP1 v1 Chiapelli McCarter
BT746889
ACCESSION BT746889
VERSION BT746889.1 GI:15768691
KEYWORDS EST.
SOURCE Meloidogyme arenaria.
ORGANISM Meloidogyme arenaria
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heterodertidae; Meloidogyninae; Meloidogyme.
REFERENCE 1 (bases 1 to 546)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pepe,D., Martin,J., Wylie,T.,
Dante,M., Maira,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvilli,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Stepien
M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James
McCarter (bchiapelli@watson.wustl.edu & jmcarter@watson.wustl.edu) at
Washington University, St. Louis. DNA sequencing by: Washington
University Genome Sequencing Center St. Louis. Nematodes were
provided by Dr. David Bird and Daniel Snyder of North Carolina

State University.
Seq primer: -40RP from GIBCO
High quality sequence stop: 439.

FEATURES

Source

1.546
/organism="Meloidogyne arenaria"
/db_xref="taxon:6304"
/clone_id="Meloidogyne arenaria egg PAMP1 v1 Chiapelli
McCarter"
/dev_str="egg"
/lab_host="DH10B"
/note="Vector: PAMP1 (Gibco); Site_1: NotI; Site_2: SalI;
The library was constructed by Brandi Chiapelli and Dr.
James McCarter at Washington University, St. Louis. The
cDNA was made by using Dynabead oligo-dT priming (Dyna).
PCR based library using a modified protocol from the
SMART PCR cDNA Synthesis Kit from Clontech. Directionally
cloned into the UDG sites of PAMP1. Nematodes were
provided by Dr. David Bird and Daniel Snyder of North
Carolina State University."

BASE COUNT

245 a 52 c 75 g 174 t

Query Match 65.0%; Score 20.8; DB 13; Length 546;
Best Local Similarity 78.1%; Pred. No. 5.4e+03;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Y 1 CAAGGTTGGAATTAAGATTAATTCGTGA 32
Db 77 CAAGGTTGGAATTAAGATTAATTCATTA 108

RESULT 11

LOCUS AQ237213 560 bp DNA linear GSS 21-APR-1999
DEFINITION RPEC11-65J23.TU RPEC1-11 Homo sapiens genomic clone RPEC1-11-65J23,
DNA sequence.

ACCESSION AQ237213
VERSION AQ237213.1 GI:3669504
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Adams, M.D., Rounsley, S.D., Zhao, S., Baas, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wilde, C., de Jong, P., and Venter, J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)

Other GSSs: RPEC11-65J23.TK
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are derived from the human BAC library RPEC1-11. For BAC
library availability, please contact Pieter de Jong
(pdejong@tigr.org, med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers
1.560
/organism="Homo sapiens"
/db_xref="GDB:752481.4"
/db_xref="taxon:9606"
/clone="RPEC1-11-65J23"
/clone_id="RPEC1-11"
/sex="Male"

Source

1.560
/organism="Homo sapiens"
/db_xref="GDB:752481.4"
/db_xref="taxon:9606"
/clone="RPEC1-11-65J23"
/clone_id="RPEC1-11"
/sex="Male"

/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPEC11 Human Male BAC library"

BASE COUNT

192 a 72 c 130 g 165 t 1 others

Query Match 65.0%; Score 20.8; DB 17; Length 560;
Best Local Similarity 78.1%; Pred. No. 5.4e+03;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Y 1 CAAGGTTGGAATTAAGATTAATTCGTGA 32
Db 370 CAAGGTTGGAATTAAGATTAATTCCTTA 401

RESULT 12

LOCUS AQ239775 662 bp DNA linear GSS 21-APR-1999
DEFINITION RPEC11-64O15.TK RPEC1-11 Homo sapiens genomic clone RPEC1-11-64O15,
DNA sequence.

ACCESSION AQ239775
VERSION AQ239775.1 GI:3668527
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Adams, M.D., Rounsley, S.D., Zhao, S., Baas, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wilde, C., de Jong, P., and Venter, J.C.
Use of human BAC End Sequences for Sequence-Ready Map Building
Unpublished (1998)

Other GSSs: RPEC11-64O15.TU
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

Clones are derived from the human BAC library RPEC1-11. For BAC
library availability, please contact Pieter de Jong
(pdejong@tigr.org, med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.

Location/Qualifiers
1.662
/organism="Homo sapiens"
/db_xref="GDB:752481.4"
/db_xref="taxon:9606"
/clone="RPEC1-11-64O15"
/clone_id="RPEC1-11"
/sex="Male"

/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPEC11 Human Male BAC Library"

BASE COUNT 220 a 109 c 122 g 211 t

Query Match 65.0%; Score 20.8; DB 17; Length 662;
Best Local Similarity 78.1%; Pred. No. 5e+03;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Y 1 CAAGGTTGGAATTAAGATTAATTCGTGA 32
Db 156 CAAGGTTGGAATTAAGATTAATTCATTA 187

RESULT 13

CNS045XL/c

LOCUS CNS045XL 669 bp DNA linear GSS 18-MAY-2000
 DEFINITION Tetradon nigroviridis genome survey sequence PUC-ori end of clone
 084H09 of library G from Tetradon nigroviridis, genomic survey
 sequence.
 ACCESSION AL275970.1 GI:8010173
 VERSION AL275970.1
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 669)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
 Saurin, W., and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetradon nigroviridis DNA sequence
 Unpublished
 JOURNAL 2 (bases 1 to 669)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
 Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetradon nigroviridis
 Unpublished
 3 (bases 1 to 669)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000)
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetradon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.
 Location/Qualifiers
 1..669
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="084H09"
 /clone_1lb="G"
 /note="Genoscope sequence ID : COBG084CD05SP1-end :
 PUC-ori"
 BASE COUNT 156 a 159 c 138 g 210 t 6 others
 ORIGIN
 Query Match 65.0%; Score 20.8; DB 17; Length 669;
 Best Local Similarity 78.1%; Pred. No. 5e+03;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 CAAGGTTCCGAATTAAGGATTAATTCGTGA 32
 Db 217 CAATGTTCTTAACGAGAGATTAATTCCTGA 186
 RESULT 14
 LOCUS AG144326 692 bp DNA linear GSS 08-JAN-2002
 DEFINITION Pan troglodytes DNA, clone: RP43-005E15.TJ, genomic survey
 sequence.
 ACCESSION AG144326
 VERSION AG144326
 KEYWORDS GSS.
 SOURCE Pan troglodytes male lymphocytes DNA, clone_1lb:RP43 Chimpanzee
 Male BAC Library clone:RP43-005E15.TJ.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 1 (bases 1 to 692)
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H., and Sakaki, Y.
 BAC end sequences of library RP43
 Unpublished

REFERENCE 2 (bases 1 to 692)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H., and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library RP43 This BAC
 end was generated during the RAD process and may have higher chance
 of clone tracking errors.
 PRIMERS
 Sequencing: TJ
 LIBRARY
 Vector : pBAC3.6
 R.Site 1 : EcoRI.
 R.Site 2 : EcoRI.
 FEATURES
 source
 1..692
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="RP43-005E15.TJ"
 /sex="male"
 /cell_type="lymphocytes"
 /clone_1lb="RP43 Chimpanzee Male BAC Library"
 BASE COUNT 229 a 133 c 125 g 205 t
 ORIGIN
 Query Match 65.0%; Score 20.8; DB 17; Length 692;
 Best Local Similarity 78.1%; Pred. No. 5e+03;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
 QY 1 CAAGGTTCCGAATTAAGGATTAATTCGTGA 32
 Db 198 CAAGGTTCTTAACGAGATTAATTCCTTA 229
 RESULT 15
 LOCUS AV853564/c 768 bp mRNA linear EST 08-NOV-2001
 DEFINITION AV853564 Nori Satoh unpublished cDNA library, larva Clona
 intestinalis cDNA clone rcilv13h17 3', mRNA sequence.
 ACCESSION AV853564
 VERSION AV853564.1 GI:166839832
 KEYWORDS EST.
 SOURCE Clona intestinalis.
 ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Clona.
 1 (bases 1 to 768)
 Satoh, N., Satou, Y., Kohara, Y., and Shin-i, T.
 Expressed genes in Clona intestinalis
 Unpublished (2000)
 JOURNAL Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 FEATURES
 source
 1..768
 /organism="Clona intestinalis"
 /db_xref="taxon:7719"
 /clone="rcilv13h17"
 /clone_1lb="Nori Satoh unpublished cDNA library, larva"
 /tissue_type="whole animal"
 /dev_stage="larva"
 /note="vector: pBluescript SK"
 BASE COUNT 212 a 138 c 174 g 242 t 2 others
 ORIGIN

Query Match 65.0%; Score 20.8; DB 10; Length 768;
 Best Local Similarity 78.1%; Pred. No. 4.8e+03;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 CAAGGTTGGAATTAAGGAATTAATTGCTGA 32
 |||||
 Db 382 CCAGAGTTCGGAATTAATCGAATTAATGCTGA 351

Search completed: May 11, 2003, 04:54:38
 Job time : 518.406 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using bw model

Run on: May 11, 2003, 03:08:16 ; Search time 31.6792 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-18

Perfect score: 32
Sequence: 1 caagggttcgaataaagaataatcgtga 32

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	100.0	32	9 US-09-963-803-18	Sequence 18, Appl
2	20	62.5	31880	9 US-10-114-170-242	Sequence 242, App
3	19.8	61.9	468	10 US-09-864-761-4932	Sequence 4932, App
4	19.4	60.6	590	10 US-09-864-761-14873	Sequence 14873, A
5	19.2	60.0	466	10 US-09-864-761-11326	Sequence 11326, A
6	19.2	60.0	38059	10 US-09-880-107-2125	Sequence 2125, App
7	19.2	60.0	684973	10 US-09-263-959-1	Sequence 1, Appl
8	18.8	58.8	435	9 US-10-101-464A-405	Sequence 405, Appl
9	18.8	58.8	9834	7 US-08-781-986A-37	Sequence 37, Appl
10	18.6	58.1	1524	10 US-09-815-242-6547	Sequence 6547, App
11	18.6	58.1	2328	9 US-09-938-842A-2679	Sequence 2679, App
12	18.6	58.1	32768	10 US-09-070-927A-128	Sequence 128, App
13	18.4	57.5	121	9 US-10-092-154-1884	Sequence 1884, App
14	18.4	57.5	121	10 US-09-764-847-1884	Sequence 1884, App
15	18.2	56.9	184	10 US-09-864-761-26009	Sequence 26009, A
16	18.2	56.9	566	10 US-09-864-761-9585	Sequence 9585, A
17	18.2	56.9	597	10 US-09-864-761-13097	Sequence 13097, A
18	18.2	56.9	764	10 US-09-770-445-942	Sequence 942, App
19	18.2	56.9	796	10 US-09-770-445-829	Sequence 829, App

20	18.2	56.9	1238	9 US-10-174-363-31	Sequence 31, Appl
21	18.2	56.9	1545	10 US-09-815-242-4389	Sequence 4389, App
22	18.2	56.9	1725	10 US-09-815-242-8228	Sequence 8228, App
23	18.2	56.9	1861	10 US-09-822-830A-4	Sequence 4, Appl
24	18.2	56.9	2102	10 US-09-764-864-268	Sequence 268, App
25	18.2	56.9	5361	9 US-09-742-096-2	Sequence 2, Appl
26	18.2	56.9	6152	9 US-09-742-096-1	Sequence 1, Appl
27	18.2	56.9	8593	9 US-09-764-904-135	Sequence 135, App
28	18.2	56.9	8593	9 US-10-091-548-135	Sequence 135, App
29	18.2	56.9	8593	9 US-10-074-095-1196	Sequence 1196, App
30	18.2	56.9	8593	10 US-09-764-860-1196	Sequence 1196, App
31	18.2	56.9	335913	9 US-09-754-853A-2	Sequence 2, Appl
32	18.2	56.9	335913	9 US-09-754-853A-3	Sequence 3, Appl
33	18.2	56.2	383	10 US-09-770-791-130	Sequence 130, App
34	18.2	56.2	433	9 US-09-918-995-8370	Sequence 8370, App
35	18.2	56.2	491	9 US-09-918-995-8666	Sequence 8666, App
36	18.2	56.2	614	10 US-09-770-149-843	Sequence 843, App
37	18.2	56.2	1020	9 US-10-121-857-69	Sequence 69, Appl
38	18.2	56.2	2000	9 US-09-938-842A-4262	Sequence 4262, App
39	18.2	56.2	11360	9 US-09-764-891-6508	Sequence 6508, App
40	18.2	56.2	11360	9 US-09-764-891-7415	Sequence 7415, App
41	18.2	56.2	55795	10 US-09-880-107-1543	Sequence 1543, App
42	18.2	56.2	640681	10 US-09-790-988-1	Sequence 1, Appl
43	17.8	55.6	284	10 US-09-560-863-726	Sequence 726, App
44	17.8	55.6	327	10 US-09-974-300-3882	Sequence 3882, App
45	17.8	55.6	406	9 US-09-918-995-4876	Sequence 4876, App

ALIGNMENTS

RESULT 1
US-09-963-803-18
Sequence 18, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 18
LENGTH: 32
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Guide desoxynucleotide building
OTHER INFORMATION: block G4
US-09-963-803-18

Query Match 100.0%; Score 32; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 CAAGGTTTCGAATAAAGGATAATTCGTGA 32
US-10-114-170-242/c
Sequence 242, Application US/10114170
Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Burland, Valerie

```

Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/114,170
FILING DATE: 01-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/453,702
FILING DATE: 03-DEC-1999
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 242:
SEQUENCE CHARACTERISTICS:
LENGTH: 31880
TYPE: nucleic acid
STRANDEDNESS: double
* TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 242:
US-10-114-170-242
Query Match 62.5%; Score 20; DB 9; Length 31880;
Best Local Similarity 82.1%; Pred. No. 3.1e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 5 GGTGGAATTAAGGAATTAATTCGTGA 32
Db 28206 GGTGTGATATGATGATTAATTTCTGA 28179
RESULT 3
US-09-864-761-4932/c
Sequence 4932, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Acomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
```

```

PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 4932
LENGTH: 468
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL031679.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.95
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.99
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.99
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.94
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
US-09-864-761-4932
Query Match 61.9%; Score 19.8; DB 10; Length 468;
Best Local Similarity 77.4%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 2 AAGGTTGGAATTAAGGAATTAATTCGTGA 32
Db 230 AATCGAGAGAAATTAAGGAATTAATGCTGA 200
RESULT 4
US-09-864-761-14873
Sequence 14873, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Acomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
```

PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14873
LENGTH: 590
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL157883.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
US-09-864-761-14873

Query Match 60.6%; Score 19.4; DB 10; Length 590;
Best Local Similarity 79.3%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 AAGGTTGGAATTAAGGATTAATTCGT 30
DB 343 AAGATTGAGAAAAAAGATTAAATTCAT 371

RESULT 5
US-09-864-761-11326/C

Sequence 11326, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11326
LENGTH: 466
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005774.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.65
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.63
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.59
US-09-864-761-11326

Query Match 60.0%; Score 19.2; DB 10; Length 466;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 CAAGGTTGGAATTAAGGATTAATTCGTCA 32
DB 188 CAAGGTTGGAATTAAGGATTAATTCCTTA 157

RESULT 6
US-09-880-107-2125

Sequence 2125, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darcie T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
PRIOR FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02

NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2125
LENGTH: 38059
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 K02402
US-09-880-107-2125

Query Match 60.0%; Score 19.2; DB 10; Length 38059;
Best Local Similarity 75.0%; Pred. No. 6.4e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CAAGGTTGGAATTAAGGATTAATTGCTGA 32
Db 27113 CAAGGTTGGAATTAAGGATTAATTGCTGA 271144

RESULT 7

US-09-263-959-1
Sequence 1, Application US/09263959
Patent No. US20020150891A1
GENERAL INFORMATION:
APPLICANT: Hood, Leroy E.
APPLICANT: Rowen, Lee
APPLICANT: Koop, Ben F.
TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTI
NUMBER OF SEQUENCES: 1279
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/263,959
FILING DATE: 05-MAR-1999
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 920010.426C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 684973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-263-959-1

Query Match 60.0%; Score 19.2; DB 10; Length 684973;
Best Local Similarity 75.0%; Pred. No. 8.4e+02;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 CAAGGTTGGAATTAAGGATTAATTGCTGA 32
Db 224824 CAAGGTTGGAATTAAGGATTAATTGCTGA 224855

RESULT 8

US-10-101-464A-405/C
Sequence 405, Application US/10101464A
Publication No. US20030046728A1

GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Nicolaas
APPLICANT: Higgins, Colleen M.
TITLE OF INVENTION: Compositions Isolated from Plant Cells
TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signaling
FILE REFERENCE: 11000.1020c2
CURRENT APPLICATION NUMBER: US/10/101,464A
CURRENT FILING DATE: 2002-03-18
PRIOR APPLICATION NUMBER: 09/704,302
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 09/228,986
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/162,866
PRIOR FILING DATE: 1999-11-01
PRIOR APPLICATION NUMBER: PCT/US00/00724
PRIOR FILING DATE: 2000-01-11
NUMBER OF SEQ ID NOS: 989
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 405
LENGTH: 435
TYPE: DNA
ORGANISM: Pinus radiata
US-10-101-464A-405

Query Match 58.8%; Score 18.8; DB 9; Length 435;
Best Local Similarity 76.7%; Pred. No. 3.8e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CAAGGTTGGAATTAAGGATTAATTGCT 30
Db 313 CAAGGATTGCCACCAAGGATTAATTCTT 284

RESULT 9

US-08-781-986A-37
Sequence 37, Application US/08781986A
Publication No. US20030054436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 9834 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
US-08-781-986A-37

Query Match
Best Local Similarity 58.8%; Score 18.8; DB 7; Length 9834;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 AGGCTCGAATAAAGATTAATTCTGTA 32
Db 5950 AGTTTAGTAAGAAGACTTAATTCTGTA 5979

RESULT 10
US-09-815-242-6547
Sequence 6547, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haebelbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PastsEQ for Windows Version 4.0
SEQ ID NO 6547
LENGTH: 1524
TYPE: DNA
ORGANISM: Enterococcus faecalis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1524)
US-09-815-242-6547

Query Match
Best Local Similarity 58.1%; Score 18.6; DB 10; Length 1524;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 TTCGAATAAAGATTAATTCTGTG 31
Db 472 TACGAATTAATGAAGAATTCTGTG 496

RESULT 11
US-09-938-842A-2679
Sequence 2679, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SRIPI3100-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 2679
LENGTH: 2328
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-2679

Query Match
Best Local Similarity 58.1%; Score 18.6; DB 9; Length 2328;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAGGTTCAATAAAGATTAATA 25
Db 1217 CAAGAGTTCAAGAAGAGATTAATA 1241

RESULT 12
US-09-070-927A-128
Sequence 128, Application US/09070927A
Patent No. US20020120116A1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
Patrick J. Dillon
Steven Barash
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 982
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 128:
SEQUENCE CHARACTERISTICS:
LENGTH: 32768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 128;
US-09-070-927A-128

Query Match 58.1%; Score 18.6; DB 10; Length 32768;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 7 TTCGAAATTAAGAAATTAATTCGTG 31
DB 19349 TACGATTAATTAAGAAATTCGTG 19373

RESULT 13
US-10-092-154-1884/C
Sequence 1884, Application US/10092154
Publication No. US20030054375A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009C1
CURRENT APPLICATION NUMBER: US/10/092.154
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2003
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1884
LENGTH: 121
TYPE: DNA
ORGANISM: Homo sapiens
US-10-092-154-1884

Query Match 57.5%; Score 18.4; DB 9; Length 121;
Best Local Similarity 78.6%; Pred. No. 4.2e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGGTTCAATTAAGAAATTAATTCGT 30
DB 115 ATGGTTCAAAATTAAGGAGTAATTCGT 88

RESULT 14
US-09-764-847-1884/C
Sequence 1884, Application US/09764847
Patent No. US20020132767A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC009
CURRENT APPLICATION NUMBER: US/09/764.847
CURRENT FILING DATE: 2001-01-17
Prior Application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 2003
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1884
LENGTH: 121
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-847-1884

Query Match 57.5%; Score 18.4; DB 10; Length 121;
Best Local Similarity 78.6%; Pred. No. 4.2e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 AGGTTCAATTAAGAAATTAATTCGT 30
DB 115 ATGGTTCAAAATTAAGGAGTAATTCGT 88

RESULT 15
US-09-864-761-26009
Sequence 26009, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Neomica-X-1
CURRENT APPLICATION NUMBER: US/09/864.761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 26009
LENGTH: 184
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC008166.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 12
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 15
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.3
OTHER INFORMATION: EST HUMAN HIT: BE077941.1, EVALUATE 7.00e-04
OTHER INFORMATION: SWISSPROT HIT: P10895, EVALUATE 1.00e+00
OTHER INFORMATION: NT HIT: AL163268.2, EVALUATE 7.00e-03
US-09-864-761-26009

Query Match 56.9%; Score 18.2; DB 10; Length 184;
Best Local Similarity 74.2%; Pred. No. 5.5e+02;
Matches 23; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 AAGGTTCAATTAAGAAATTAATTCGTGA 32
DB 143 ATGGATTGAAAGAAAGAAATTAATTCGA 173

Mon May 12 14:15:16 2003

Search completed: May 11, 2003, 06:32:41
Job time : 110.679 secs

us-09-963-803-18.rmpb

Page 7

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 688.446 Seconds
(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65
Sequence: 1 cagctgcagactatcc.....tactgagagatgatagct 65

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: 1: gb_ba: 2: gb_hg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_stb: 12: gb_sy: 13: gb_un: 14: gb_vi: 15: em_ba: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_stb: 28: em_un: 29: em_vi: 30: em_hg_hum: 31: em_hg_inv: 32: em_hg_other: 33: em_hg_mus: 34: em_hg_pln: 35: em_hg_rtd: 36: em_hg_mam: 37: em_hg_vrt: 38: em_sy: 39: em_hggo_hum: 40: em_hggo_mus: 41: em_hggo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	100.0	65	AX036742	AX036742 Sequence
2	65	100.0	317	AX036737	AX036737 Sequence
3	65	100.0	371	AX036739	AX036739 Sequence
4	58	89.2	348	AX036738	AX036738 Sequence
5	58	89.2	398	AX036740	AX036740 Sequence
6	58	89.2	472	AX036759	AX036759 Sequence
7	58	89.2	541	AX036758	AX036758 Sequence
8	58	89.2	604	AX036757	AX036757 Sequence
9	55	84.6	332	AX036755	AX036755 Sequence
10	55	84.6	393	AX036753	AX036753 Sequence
11	55	84.6	462	AX036754	AX036754 Sequence
12	55	84.6	600	AX036756	AX036756 Sequence
13	55	84.6	9285	AX036754	AX036754 Sequence
14	55	84.6	15077	AX036754	AX036754 Sequence
15	47	72.3	7489	AX036754	AX036754 Sequence
16	46	70.8	243	AX036735	AX036735 Sequence
17	28.8	44.3	1381	AX036735	AX036735 Sequence
18	28.8	43.4	81480	AX036735	AX036735 Sequence
19	27.8	42.8	168367	AX036735	AX036735 Sequence
20	27.4	42.2	214175	AX036735	AX036735 Sequence
21	27.2	41.8	16357	AX036735	AX036735 Sequence
22	27.2	41.8	169862	AX036735	AX036735 Sequence
23	27.2	41.8	172346	AX036735	AX036735 Sequence
24	26.8	41.2	142559	AX036735	AX036735 Sequence
25	26.8	41.2	145535	AX036735	AX036735 Sequence
26	26.8	41.2	148792	AX036735	AX036735 Sequence
27	26.8	41.2	149055	AX036735	AX036735 Sequence
28	26.8	41.2	164564	AX036735	AX036735 Sequence
29	26.8	41.2	178814	AX036735	AX036735 Sequence
30	26.8	41.2	186491	AX036735	AX036735 Sequence
31	26.8	41.2	195240	AX036735	AX036735 Sequence
32	26.8	41.2	196877	AX036735	AX036735 Sequence
33	26.8	41.2	242569	AX036735	AX036735 Sequence
34	26.6	40.9	128250	AX036735	AX036735 Sequence
35	26.6	40.9	146024	AX036735	AX036735 Sequence
36	26.6	40.9	177514	AX036735	AX036735 Sequence
37	26.6	40.9	194418	AX036735	AX036735 Sequence
38	26.4	40.6	50979	AX036735	AX036735 Sequence
39	26.4	40.6	164719	AX036735	AX036735 Sequence
40	26.4	40.6	335871	AX036735	AX036735 Sequence
41	26.2	40.3	636	AX036735	AX036735 Sequence
42	26.2	40.3	156764	AX036735	AX036735 Sequence
43	26	40.0	2566	AX036735	AX036735 Sequence
44	26	40.0	104383	AX036735	AX036735 Sequence
45	26	40.0	138556	AX036735	AX036735 Sequence

ALIGNMENTS

RESULT 1
AX036742
LOCUS
DEFINITION
ACCESSION
VERSION
SOURCE
KEYWORDS
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AX036742
Sequence 8 from Patent WO0058485.
AX036742
AX036742.1 GI:11226251
synthetic construct.
synthetic construct.
artificial sequences.
1 (bases 1 to 65)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 8 05-OCT-2000;
JOURNAL

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
SOURCE
1. 65
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional desoxynucleotide building block S1"
BASE COUNT 20 a 16 c 14 g 15 t
ORIGIN

Query Match 100.0%; Score 65; DB 6; Length 65;
Best Local Similarity 100.0%; Pred. No. 5.2e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGCTGACAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 60
Db 1 CATGCTGACAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 60
Qy 61 TAGCT 65
Db 61 TAGCT 65

RESULT 2
AX036737 317 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 3 from Patent WO0058485.
ACCESSION AX036737
VERSION AX036737.1 GI:11226246
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 317)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 3 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
SOURCE
Location/Qualifiers
1. .317
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter MP1116"
BASE COUNT 107 a 61 c 74 g 75 t
ORIGIN

Promoter 1. .317
Query Match 100.0%; Score 65; DB 6; Length 317;
Best Local Similarity 100.0%; Pred. No. 6.1e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGCTGACAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 60
Db 8 CATGCTGACAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 67
Qy 61 TAGCT 65
Db 68 TAGCT 72

RESULT 3
AX036739 371 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 5 from Patent WO0058485.
ACCESSION AX036739
VERSION AX036739.1 GI:11226248
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 371)

Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 5 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
SOURCE
1. .371
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter MP1116"
BASE COUNT 122 a 68 c 89 g 92 t
ORIGIN

Promoter 1. .371
Query Match 100.0%; Score 65; DB 6; Length 371;
Best Local Similarity 100.0%; Pred. No. 6.2e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CATGCTGACAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 60
Db 8 CATGCTGACAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAA 67
Qy 61 TAGCT 65
Db 68 TAGCT 72

RESULT 4
AX036738 348 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 4 from Patent WO0058485.
ACCESSION AX036738
VERSION AX036738.1 GI:11226247
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 348)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 4 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
SOURCE
Location/Qualifiers
1. .348
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter MP1117"
BASE COUNT 116 a 70 c 78 g 84 t
ORIGIN

Promoter 1. .348
Query Match 89.2%; Score 58; DB 6; Length 348;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 CAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAAAGTACT 65
Db 49 CAGACTAGTATCCGCGTCATCATGACATCATCAGTACTGAGAGATGAAATAGCT 106

RESULT 5
AX036740 398 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 6 from Patent WO0058485.
ACCESSION AX036740
VERSION AX036740.1 GI:11226249
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.

REFERENCE 1 (bases 1 to 398)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 6 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. 398
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1147"
BASE COUNT promoter 128 a 80 c 93 g 97 t
ORIGIN
Query Match 89.2%; Score 58; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 49 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 106
RESULT 6
LOCUS AX036759 472 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 25 from Patent WO0058485.
ACCESSION AX036759
VERSION AX036759.1 GI:11226268
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 472)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 25 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. 472
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1169"
BASE COUNT promoter 149 a 92 c 112 g 119 t
ORIGIN
Query Match 89.2%; Score 58; DB 6; Length 472;
Best Local Similarity 100.0%; Pred. No. 4.1e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 49 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 106
RESULT 7
LOCUS AX036758 541 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 24 from Patent WO0058485.
ACCESSION AX036758
VERSION AX036758.1 GI:11226267
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 541)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow

JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 24 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. 541
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1168"
BASE COUNT promoter 169 a 104 c 130 g 138 t
ORIGIN
Query Match 89.2%; Score 58; DB 6; Length 541;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 49 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 106
RESULT 8
LOCUS AX036757 604 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 23 from Patent WO0058485.
ACCESSION AX036757
VERSION AX036757.1 GI:11226266
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 604)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 23 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. 604
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1167"
BASE COUNT promoter 186 a 116 c 145 g 157 t
ORIGIN
Query Match 89.2%; Score 58; DB 6; Length 604;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 49 CAGACTAGTATCCGCGCTCATCAATGACATCATCAGTACTGAGAGATGAATAGCT 106
RESULT 9
LOCUS AX036755 392 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 21 from Patent WO0058485.
ACCESSION AX036755
VERSION AX036755.1 GI:11226264
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 392)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 21 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

WMRFKDEYONLAKIANQDGTQAILSQIRRIFLGEDPVLGONTVONIAPRKLOLVC
PNVOSIRRYLMDYLTTLAETGLMSETEGPALISELFTKMPAIGERVAOAYKIMPT
SAVILPSRYFTIINYLTAEQCEASMYRSKALDECRDPFIEGYGRSEKKYATARA
TKYTKGADHDIRTAKYORKCKCYI CGOGEHYANOCRNKHQOORVALILOSILKE
NEEVSADDEEEDDELFSVLGEDYOEETIMWIEEDIOOIKEFSKRGDLSRRNG
PNFPGAEVQGVLPKPSWRRIQATLEBINCHNMTAISTQLACRSCKQPLAGVO
CHCHAVYCFMCAEAYHDVOAKILSKDYSPSARGKGAVIIIEDEIBGEPLISOLQ
OENORLOKQVRLQOEELMLKREKDEALKHSEKASRYFTIOESDEALNLIKELRO
FKEETRMALIQLEKALIVOEEDTIEERCAMILEEKHTENIYSATAKAENGLYVKG
IKPDNMEPYIINAVDGTATACLIQISAI PENNYEDAKVTVPNSVIGISTOMIKA
GRILLIGQYFRMPTYIMNGLSPGIOMIIGCSFIRLSEGLRIEKDITTPYKLTISI
ETSRITOVANSIELELSEDEYINLASVETPSFLDOEFARKNDLLEKEMKYEIGE
NPMPEWNNKIKCCLNIINPDIKIMGRPIKGVTPGDEAMTROINLLQMKVIRPSES
KHSSTAFIVRSGETIDPIITGKEKKERWENYGLINENTESDQYSLPGINTIISKVG
RSKIYSEKDLKSGFMOVAMESESVPVTAFLAGNGLYEMLVMPGLKNAPAIFORXDMN
VPFGTEKEIYAVYIDDIIVFSETAQHSOHLVTMQLCKENGILISPTMKIGTPEIDF
LGASLGCTKIKLOPHITISKICDPSDEKLTATPEGRSWGLISTARNTYODIGKLOPL
ROKAPFTGDYKRNPEETWKNVQIKERKVNLPDLQLPKDSFIIIEYDGMWGAVCK
WMSKIDPRSTERICAYASGSFNPDKSTIDAEIOAIIHGLDKPKIYVLDKRELIIRSD
CEALIKFYNNTNENKPSRVWLFPSPDLTGLGITVTEPHIDGKHNGIADALSRMINPI
VERKDESPYRFTSSVEDALKVCNDHGRNLISAVINDIITVLR

variation

variation

old_sequence

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

variation

BASE COUNT 2662 a 1389 c 1577 g 1861 t
ORIGIN

Query Match 72.3%; Score 47; DB 14; Length 7489;

Best Local Similarity 100.0%; Pred. No. 1.4e-07;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TATCCGCGCATCATGATCATGACATGACTGAGAGATGAATA 62
DB 7123 TATCCGCGCATCATGATCATGACATGACTGAGAGATGAATA 7169

Search completed: May 11, 2003, 06:04:51
Job time : 695.446 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 126.416 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65
Sequence: 1 catgctgcagactatcc.....tactgagagatcatagct 65

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseg-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	100.0	65	21	AAA96842
2	65	100.0	317	21	AAA96837
3	65	100.0	371	21	AAA96839
4	58	89.2	348	21	AAA96838
5	58	89.2	398	21	AAA96840
6	58	89.2	472	21	AAA96859
7	58	89.2	541	21	AAA96858
8	58	89.2	604	21	AAA96857
9	55	84.6	392	21	AAA96855

10	55	84.6	393	21	AAA96853	Nucleotide sequenc
11	55	84.6	462	21	AAA96854	Nucleotide sequenc
12	55	84.6	600	21	AAA96856	Nucleotide sequenc
13	46	70.8	243	21	AAA96835	Promoter from inte
14	26.4	40.6	2837	23	AB111216	Drosophila melanog
15	24.8	38.2	5668	23	AB114687	Drosophila melanog
16	24.8	38.2	9477	23	AB114686	Drosophila melanog
17	24.4	37.5	1006	21	AA133441	Arabidopsis thalia
18	24.2	37.2	2600	17	AA133624	Chromatin regulato
19	24.2	37.2	2600	14	AB165847	Lung cancer relate
20	24.2	37.2	22798	22	AA181124	Human immune/haema
21	24.2	37.2	22798	22	AA181124	Human immune/haema
22	24	36.9	452	23	AA184412	DNA encoding novel
23	24	36.9	569	23	AA184412	Human prostate exp
24	24	36.9	755	22	AA175159	Nucleotide sequenc
25	24	36.9	755	22	AA175159	Human polynucleoti
26	24	36.9	2320	21	AA177735	Arabidopsis thalia
27	23.8	36.6	1467	23	AA184414	DNA encoding novel
28	23.8	36.6	30115	21	AA184414	Modified adenovira
29	23.6	36.3	273	22	AA184051	Human ovarian PCR-
30	23.6	36.3	385	23	AA182620	Human ovarian tumo
31	23.6	36.3	1863	23	AB121739	Drosophila melanog
32	23.6	36.3	2884	23	AB121738	Drosophila melanog
33	23.6	36.3	4168	24	AB192106	Human Tumour Endot
34	23.4	36.0	424	22	AB114122	Human breast cance
35	23.4	36.0	468	22	AA112735	Human breast cance
36	23.4	36.0	570	22	AA122984	Human breast cance
37	23.4	36.0	764	22	AA121607	Human breast cance
38	23.4	36.0	147724	22	AB183566	Human CDNA differe
39	23.2	35.7	470	22	AA131993	Human olfactory re
40	23	35.4	1809	21	AA131993	Human olfactory re
41	22.8	35.1	480	21	AA131993	Human secreted pro
42	22.8	35.1	349980	22	AA131993	Eucalyptus grandis
43	22.6	34.8	459	22	AB144836	Pyrococcus abyssi
44	22.6	34.8	459	22	AB144836	Human breast cell
45	22.6	34.8	459	22	AB144836	Human foetal liver

ALIGNMENTS

RESULT 1	AAA96842	standard; DNA; 65 BP.
AAA96842	standard; DNA; 65 BP.	
AC	AAA96842:	
XX		
DT	19-FEB-2001	(first entry)
XX		
DE	Directional desoxynucleotide building block SL.	
XX		
KM	Promoter, intergenic region; Comelina yellow mottle virus;	
KM	Chimeric expression promoter; plant vascular expression promoter;	
KM	plant green tissue expression promoter; Cassava vein mosaic virus;	
KM	transgenic plant; ss.	
XX		
OS	Synthetic.	
XX		
PN	WO200058485-A1.	
XX		
PD	05-OCT-2000.	
XX		
PF	29-MAR-2000; 2000WO-IB00370.	
XX		
PR	29-MAR-1999; 99FR-0003925.	
XX		
PA	(MERI-) MERISTEM THERAPEUTICS.	
XX		
PI	Rance I, Gruber V, Theisen M;	
XX		
DR	WPI; 2000-647238/62.	
XX		
PT	Chimeric expression promoter for transgenic plant production, comprises	

PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX Disclosure; Page 23; 91pp; English.
XX
CC The present sequence represents a directional deoxynucleotide building
CC block, which was used to construct chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comellina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

SO Sequence 65 BP; 20 A; 16 C; 14 G; 15 T; 0 other;

Query Match 100.0%; Score 65; DB 21; Length 65;

Best Local Similarity 100.0%; Pred. No. 5.1e-15;

Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGCTGAGACTAGTATCCGCGTCATCATGACATCATGAGAGATGAA 60

DB 1 CATGCTGAGACTAGTATCCGCGTCATCATGACATCATGAGAGATGAA 60

QY 61 TAGCT 65

DB 61 TAGCT 65

RESULT 2

AAA96837 standard; DNA; 317 BP.

AC AAA96837;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1116.

KM Promoter; intergenic region; Comellina yellow mottle virus;

KW chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

OS Chimeric - Comellina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN W0200058485-A1.

PF 29-MAR-2000; 2000MO-IB00370.

PR 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

PI Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 81; 91pp; English.

CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comellina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

SO Sequence 317 BP; 107 A; 61 C; 74 G; 75 T; 0 other;

Query Match 100.0%; Score 65; DB 21; Length 317;

Best Local Similarity 100.0%; Pred. No. 8e-15; Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGCTGAGACTAGTATCCGCGTCATCATGACATCATGAGAGATGAA 60

DB 8 CATGCTGAGACTAGTATCCGCGTCATCATGACATCATGAGAGATGAA 67

QY 61 TAGCT 65

DB 68 TAGCT 72

RESULT 3

AAA96839 standard; DNA; 371 BP.

AC AAA96839;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1116.

KM Promoter; intergenic region; Comellina yellow mottle virus;

KW chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

OS Chimeric - Comellina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN W0200058485-A1.

PF 29-MAR-2000; 2000MO-IB00370.

PR 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

PI Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 81; 91pp; English.

CC The present sequence represents a chimeric promoter of the invention.

CC The specification describes chimeric expression promoters. These

CC chimeric promoters comprise a nucleic acid sequence which is derived

CC from a first plant promoter, in which a plant vascular expression

CC promoter region is replaced with a nucleic acid sequence derived from

CC a second plant promoter comprising a plant green tissue expression

CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX Sequence 371 BP; 122 A; 68 C; 89 G; 92 T; 0 other;

Query Match 100.0%; Score 65; DB 21; Length 371;
Best Local Similarity 100.0%; Pred. No. 8.3e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATCTGACAGTATGATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAA 60
DB 8 CATCTGACAGTATGATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAA 67

QY 61 TAGCT 65
DB 68 TAGCT 72

RESULT 4

AAA96838 standard; DNA; 348 BP.

AC AAA96838;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1117.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KM plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

XX Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

PD 05-OCT-2000.

XX 29-MAR-2000; 2000WO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

XX (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

XX WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

XX sequence from promoter comprising vascular expression region replaced

XX with sequence from promoter comprising green tissue expression region

XX Claim 5; Page 81; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

XX The specification describes chimeric expression promoters. These

XX chimeric promoters comprise a nucleic acid sequence which is derived

XX from a first plant promoter, in which a plant vascular expression

XX promoter region is replaced with a nucleic acid sequence derived from

XX a second plant promoter comprising a plant green tissue expression

XX promoter region. Preferably, the first plant promoter originates from

XX Commelina yellow mottle virus, and the second plant promoter originates

XX from the Cassava vein mosaic virus. Especially, the promoters are

XX derived from intergenic regions. The chimeric promoters are useful

XX for producing transgenic plants.

Query Match 89.2%; Score 58; DB 21; Length 348;
Best Local Similarity 100.0%; Pred. No. 3.6e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTGATATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAAATAGCT 65
DB 49 CAGACTGATATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAAATAGCT 106

RESULT 5

AAA96840 standard; DNA; 398 BP.

AC AAA96840;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1147.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KM plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

XX Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

PD 05-OCT-2000.

XX 29-MAR-2000; 2000WO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

XX (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

XX WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

XX sequence from promoter comprising vascular expression region replaced

XX with sequence from promoter comprising green tissue expression region

XX Claim 5; Page 82; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

XX The specification describes chimeric expression promoters. These

XX chimeric promoters comprise a nucleic acid sequence which is derived

XX from a first plant promoter, in which a plant vascular expression

XX promoter region is replaced with a nucleic acid sequence derived from

XX a second plant promoter comprising a plant green tissue expression

XX promoter region. Preferably, the first plant promoter originates from

XX Commelina yellow mottle virus, and the second plant promoter originates

XX from the Cassava vein mosaic virus. Especially, the promoters are

XX derived from intergenic regions. The chimeric promoters are useful

XX for producing transgenic plants.

Query Match 89.2%; Score 58; DB 21; Length 398;
Best Local Similarity 100.0%; Pred. No. 3.7e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTGATATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAAATAGCT 65
DB 49 CAGACTGATATCCGCGTCATCAATGACATCATCACTAGTGGAGATGAAATAGCT 106

RESULT 6

AAA96859

ID	AA96859 standard; DNA; 472 BP.
XX	
AC	AA96859;
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Nucleotide sequence of chimeric expression promoter MP1169.
XX	
KW	Promoter; intergenic region; Commelina yellow mottle virus;
KW	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
XX	transgenic plant; chimera; ss.
XX	
OS	Chimeric - Commelina yellow mottle virus.
OS	Chimeric - Cassava vein mosaic virus.
XX	
PN	MO200058485-A1.
XX	
PD	05-OCT-2000.
XX	
PF	29-MAR-2000; 2000MO-1B00370.
XX	
PR	29-MAR-1999; 99PR-0003925.
XX	
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	
PI	Rance I, Gruber V, Theisen M;
XX	
DR	WPI; 2000-647238/62.
PT	
PT	Chimeric expression promoter for transgenic plant production, comprises
PT	sequence from promoter comprising vascular expression region replaced
PT	with sequence from promoter comprising green tissue expression region
PT	
PS	
PS	Claim 5; Page 88; 91pp; English.
XX	
CC	The present sequence represents a chimeric promoter of the invention.
CC	The specification describes chimeric expression promoters. These
CC	chimeric promoters comprise a nucleic acid sequence which is derived
CC	from a first plant promoter, in which a plant vascular expression
CC	promoter region is replaced with a nucleic acid sequence derived from
CC	a second plant promoter comprising a plant green tissue expression
CC	promoter region. Preferably, the first plant promoter originates from
CC	Commelina yellow mottle virus, and the second plant promoter originates
CC	from the Cassava vein mosaic virus. Especially, the promoters are
CC	derived from intergenic regions. The chimeric promoters are useful
CC	for producing transgenic plants.
XX	
SO	Sequence 472 BP; 149 A; 92 C; 112 G; 119 T; 0 other;
SO	
Query Match	89.2%; Score 58; DB 21; Length 472;
Best Local Similarity	100.0%; Pred. No. 3.9e-12;
Matches	58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	8 CAGACTAGTATCCGCCGTCATCATGACATCATCACAGTCTGAGAGATGAATGACT 65
DB	49 CAGACTAGTATCCGCCGTCATCATGACATCATCACAGTCTGAGAGATGAATGACT 106
RESULT 7	
ID	AA96858
XX	AA96858 standard; DNA; 541 BP.
XX	
AC	AA96858;
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Nucleotide sequence of chimeric expression promoter MP1168.
XX	
KW	Promoter; intergenic region; Commelina yellow mottle virus;
KW	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;

KW		transgenic plant; chimera; ss.
OS		Chimeric - Commelina yellow mottle virus.
OS		Chimeric - Cassava vein mosaic virus.
XX		
PN		MO200058485-A1.
XX		
PD		05-OCT-2000.
XX		
XP		29-MAR-2000; 2000WO-IB00370.
XX		
XX		
PR		29-MAR-1999; 99FR-0003925.
XX		
PA		(MERI-) MERISTEM THERAPEUTICS.
XX		
P1		Rance I, Gruber V, Theisen M;
XX		
DR		WPI; 2000-647238/62.
XX		
PT		Chimeric expression promoter for transgenic plant production, comprises
PT		sequence from promoter comprising vascular expression region replaced
PT		with sequence from promoter comprising green tissue expression region
PT		-
PS		Claim 5; Page 87-88; 91pp; English.
CC		The present sequence represents a chimeric promoter of the invention.
CC		The specification describes chimeric expression promoters. These
CC		chimeric promoters comprise a nucleic acid sequence which is derived
CC		from a first plant promoter, in which a plant vascular expression
CC		promoter region is replaced with a nucleic acid sequence derived from
CC		a second plant promoter comprising a plant green tissue expression
CC		promoter region. Preferably, the first plant promoter originates from
CC		Commelina yellow mottle virus, and the second plant promoter originates
CC		from the Cassava vein mosaic virus. Especially, the promoters are
CC		derived from intergenic regions. The chimeric promoters are useful
CC		for producing transgenic plants.
SQ		Sequence 541 BP, 169 A, 104 C, 130 G, 138 T, 0 other;
		Query Match 89.2%; Score 58; DB 21; Length 541;
		Best Local Similarity 100.0%; Pred. No. 4e-12;
		Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY		
		8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTAGTGAGAGATAATGCT 65
Dd		49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGTAGTGAGAGATAATGCT 106
		RESULT 8
		AAA96857
XX	ID	AAA96857 standard; DNA; 604 BP.
XX	AC	AAA96857;
XX	DT	19-FEB-2001 (first entry)
XX	DE	Nucleotide sequence of chimeric expression promoter MPRI167.
XX		
KW		Promoter; intergenic region; Commelina yellow mottle virus;
KW		chimeric expression promoter; plant vascular expression promoter;
KW		plant green tissue expression promoter; Cassava vein mosaic virus;
KW		transgenic plant; chimera; ss.
XX		
OS		Chimeric - Commelina yellow mottle virus.
OS		Chimeric - Cassava vein mosaic virus.
XX		
FN		MO200058485-A1.
XX		
PD		05-OCT-2000.
XX		
XP		29-MAR-2000; 2000WO-IB00370.
XX		

PR 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
PI Rance I, Gruber V, Theisen M;
XX WPI; 2000-647238/62.
DR
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
XX Claim 5; Page 87; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 604 BP; 186 A; 116 C; 145 G; 157 T; 0 other;
XX
Query Match 89.2%; Score 58; DB 21; Length 604;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 8 CAGCTGATCCCGCTCATCATGACATCATCATGACTACTGAGGATGAATGCT 65
DB 49 CAGCTGATCCCGCTCATCATGACATCATCATGACTACTGAGGATGAATGCT 106
XX
RESULT 9
ID AAA96855 standard; DNA; 392 BP.
XX
AC AAA96855;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter Mpr1164.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
XX
XX 29-MAR-2000; 2000WO-IB00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT

PT -
XX
XX Claim 5; Page 86; 91pp; English.
PS
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 392 BP; 127 A; 80 C; 87 G; 98 T; 0 other;
XX
Query Match 84.6%; Score 55; DB 21; Length 392;
Best Local Similarity 100.0%; Pred. No. 5e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 11 ACTGATATCCCGCTCATCATGACATCATCATGACTACTGAGGATGAATGCT 65
DB 20 ACTGATATCCCGCTCATCATGACATCATCATGACTACTGAGGATGAATGCT 74
XX
RESULT 10
ID AAA96853 standard; DNA; 393 BP.
XX
AC AAA96853;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter Mpr1162.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
XX
XX 29-MAR-2000; 2000WO-IB00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
XX Claim 5; Page 85; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

SO Sequence 393 BP, 128 A, 75 C, 93 G, 97 T, 0 other;

Query Match 84.6%; Score 55; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 5e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ACTGATCCCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 65
DB 20 ACTGATATCCCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 74

RESULT 11

AAA96854
ID AAA96854 standard; DNA; 462 BP.

AC AAA96854;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MPr1163.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KM plant green tissue expression promoter; Cassava vein mosaic virus;

KM transgenic plant; chimera; 88.

XX Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

XX WO200058485-A1.

PN 05-OCT-2000.

PF 29-MAR-2000; 2000MO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 86; 91pp; English.

CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the first plant promoter originates from
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX Sequence 462 BP, 148 A, 87 C, 111 G, 116 T, 0 other;

Query Match 84.6%; Score 55; DB 21; Length 462;
Best Local Similarity 100.0%; Pred. No. 5.2e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ACTGATCCCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 65
DB 20 ACTGATATCCCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 74

RESULT 12

AAA96856
ID AAA96856 standard; DNA; 600 BP.

AC AAA96856;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MPr165.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KM chimeric expression promoter; plant vascular expression promoter;

KM plant green tissue expression promoter; Cassava vein mosaic virus;

KM transgenic plant; chimera; 88.

XX Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

PD 05-OCT-2000.

PF 29-MAR-2000; 2000MO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 86-87; 91pp; English.

CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX Sequence 600 BP, 188 A, 111 C, 147 G, 154 T, 0 other;

Query Match 84.6%; Score 55; DB 21; Length 600;
Best Local Similarity 100.0%; Pred. No. 5.6e-11;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 ACTGATCCCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 65
DB 20 ACTGATATCCCGCTCATCATGACATCATCAGTACTGAGAGATGATAGCT 74

RESULT 13

AAA96835
ID AAA96835 standard; DNA; 243 BP.

AC AAA96835;

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX PS Claim 1; SEQ ID NO 38543; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 5668 BP; 1478 A; 1411 C; 1420 G; 1359 T; 0 other;

Query Match 38.2%; Score 24.8; DB 23; Length 5668;

Best Local Similarity 63.3%; Pred. No. 26;

Matches 38; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 1 CATCTGCAGACTAGTATCCGCGTCATCAATGACATCATCAGTACTGAGAGATGAA 60

DB 667 CATCTGCTTCATGATGCGCTTCATCTTCGACTTCATCTCAATGTTGAAGATCTGCA 608

Search completed: May 11, 2003, 03:04:00
Job time : 128.416 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: May 11, 2003, 02:44:10 ; Search time 26.0652 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65
Sequence: 1 catgctgcagactatcc.....tactgagagatgatagct 65

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22.2	34.2	1086	US-09-183-861-77	Sequence 77, Appl
C 2	22.2	34.2	1086	US-09-022-765-77	Sequence 77, Appl
C 3	22.2	34.2	1179	US-09-147-926-1	Sequence 1, Appl
C 4	22.2	33.8	4010	US-08-785-310A-3	Sequence 3, Appl
C 5	21.8	33.5	1515	US-09-134-001C-143	Sequence 143, App
C 6	21.8	33.5	5393	US-08-591-079-9	Sequence 9, Appl
C 7	21.8	33.5	13473	PCT-US96-03916-1	Sequence 1, Appl
C 8	21.8	33.5	18912	PCT-US96-03916-59	Sequence 59, Appl
C 9	21.6	33.2	4940	US-08-484-105-1	Sequence 1, Appl
C 10	21.6	33.2	4940	US-08-484-106-1	Sequence 1, Appl
C 11	21.6	33.2	8561	US-09-112-450-3	Sequence 3, Appl
C 12	21.6	33.2	8561	US-09-419-281A-3	Sequence 3, Appl
C 13	21.6	32.6	1001	US-09-641-638-111	Sequence 111, App
C 14	21.2	32.6	1011	US-09-641-638-112	Sequence 112, App
C 15	21.2	32.6	3116	US-09-362-831-10	Sequence 10, Appl
C 16	21.2	32.6	7082	US-09-362-831-1	Sequence 1, Appl
C 17	21.2	32.6	1525	US-08-186-833-3	Sequence 3, Appl
C 18	21.2	32.3	2085	US-09-289-843A-65	Sequence 65, Appl
C 19	21.2	32.3	2085	US-09-088-357B-65	Sequence 65, Appl
C 20	21.2	32.3	4403765	US-09-103-840A-2	Sequence 2, Appl
C 21	21.2	32.3	4411529	US-09-103-840A-1	Sequence 1, Appl
C 22	20.8	32.0	1010	US-09-453-702B-4	Sequence 4, Appl
C 23	20.8	32.0	2818	US-08-366-276-1	Sequence 1, Appl
C 24	20.8	32.0	3282	US-08-072-574-11	Sequence 11, Appl
C 25	20.8	32.0	3282	US-08-486-270-11	Sequence 11, Appl
C 26	20.8	32.0	3282	US-08-367-264-11	Sequence 11, Appl
C 27	20.8	32.0	3282	US-09-153-757-11	Sequence 11, Appl

C 28	20.8	32.0	4085	US-08-072-574-7	Sequence 7, Appl
C 29	20.8	32.0	4085	US-08-486-270-7	Sequence 7, Appl
C 30	20.8	32.0	4085	US-08-367-264-7	Sequence 7, Appl
C 31	20.8	32.0	4085	US-09-153-757-7	Sequence 7, Appl
C 32	20.8	32.0	4181	US-08-072-574-9	Sequence 9, Appl
C 33	20.8	32.0	4181	US-08-486-270-9	Sequence 9, Appl
C 34	20.8	32.0	4181	US-08-367-264-9	Sequence 9, Appl
C 35	20.8	32.0	4181	US-09-153-757-9	Sequence 9, Appl
C 36	20.8	32.0	4207	US-08-660-148-1	Sequence 1, Appl
C 37	20.8	32.0	4207	US-08-660-148-3	Sequence 3, Appl
C 38	20.8	32.0	4303	US-08-660-148-4	Sequence 4, Appl
C 39	20.8	32.0	4303	US-08-660-148-6	Sequence 6, Appl
C 40	20.8	32.0	7096	US-09-221-017B-373	Sequence 373, App
C 41	20.8	32.0	8257	US-09-484-970B-65	Sequence 65, Appl
C 42	20.8	32.0	46819	US-09-453-702B-72	Sequence 72, Appl
C 43	20.6	31.7	1080	US-09-147-926-3	Sequence 3, Appl
C 44	20.6	31.7	9179	US-09-453-702B-162	Sequence 162, App
C 45	20.4	31.4	463	US-08-943-731-82	Sequence 82, Appl

ALIGNMENTS

RESULT 1
US-09-183-861-77/c
Sequence 77, Application US/09183861
Patent No. 6365165
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, David C.
APPLICANT: Skelky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09183,861
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/022,765
FILING DATE: 12-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 77:
SEQUENCE CHARACTERISTICS:
LENGTH: 1086 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-183-861-77

Query Match 34.2%; Score 22.2; DB 4; Length 1086;
Best Local Similarity 59.0%; Pred. No. 22;
Matches 36; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 3 TGCTGCACTTAGTATCCGCGCTCATCATATACATCATCAGTAGTGGAGATGATA 62
Db 246 TGTAGCAGCGTCTCAGCAGCCGGGAGAAATTATCATCATACGCTTTCAGAACTTTTC 187
QY 63 G 63
Db 186 G 186

RESULT 2
US-09-022-765-77/c

Query Match	34.2%	Score 22.2	DB 4	Length 1086
Best Local Similarity	59.0%	Pred. No. 22		
Matches 36	Conservative 0	Mismatches 25	Indels 0	Gaps 0

RESULT 3
US-09-147-926-1/C
Sequence 1, Application US/09147926
Patent No. 6197549
GENERAL INFORMATION
APPLICANT: Palmer, Leslie M.
APPLICANT: Pratt, Julie W.
APPLICANT: Hodgson, John E.

```

1  APPLICANT: Beattie, David T.
2  APPLICANT: Lowe, Adrian M.
3  APPLICANT: Lometto, Michael A.
4  APPLICANT: Nichols, Richard O.
5  APPLICANT: Detsilewicz, Robert L.
6  TITLE OF INVENTION: ama
7  FILE REFERENCE: CM10107
8  CURRENT APPLICATION NUMBER: US/09/147,926
9  CURRENT FILING DATE: 1999-03-19
10 PRIOR APPLICATION NUMBER: 60/061,077
11 PRIOR FILING DATE: 1997-10-03
12 NUMBER OF SEQ ID NOS: 6
13 SOFTWARE: FASTSEQ for Windows Version 3.
14 SEQ ID NO 1
15     LENGTH: 1179
16     TYPE: DNA
17     ORGANISM: Staphylococcus aureus
18     FEATURE:
19     NAME/KEY: CDS
20     LOCATION: (1) ... (1176)
21 US-09-147-926-1

```

Query Match	34.2%	Score 22.2;	DB 4;	Length 1179;
Best Local Similarity	61.0%	Pred. No. 22;		
Matches 36; Conservative	0;	Mismatches 23;	Indels 0;	Gaps 0;

```

1      RESULT 4
2      US-08-785-310A-3
3      Sequence 3, Application US/08785310A
4      Patent No. 5840532
5      GENERAL INFORMATION:
6      APPLICANT: McKnight, Steven L.
7      APPLICANT: Russell, David W.
8      TITLE OF INVENTION: Neuronal PAS Domain Protein
9      NUMBER OF SEQUENCES: 8
10     CORRESPONDENCE ADDRESS:
11     ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
12     STREET: 268 BUSH STREET, SUITE 3200
13     CITY: SAN FRANCISCO
14     STATE: CALIFORNIA
15     COUNTRY: USA
16     ZIP: 94104
17     COMPUTER READABLE FORM:
18     MEDIUM TYPE: Floppy disk
19     COMPUTER: IBM PC compatible
20     OPERATING SYSTEM: PC-DOS/MS-DOS
21     SOFTWARE: PatentIn Release #1.0, Version #1.3JGD
22     CURRENT APPLICATION DATA:
23     APPLICATION NUMBER: US/08/785,310A
24     FILING DATE: 21-JAN-1997
25     CLASSIFICATION: 536
26     ATTORNEY/AGENT INFORMATION:
27     NAME: OSMAN, RICHARD A
28     REGISTRATION NUMBER: 36,627
29     REFERENCE/DOCKET NUMBER: UTSD:1226
30     TELECOMMUNICATION INFORMATION:
31     TELEPHONE: (415) 343-4341
32     TELEFAX: (415) 343-4342
33     INFORMATION FOR SEQ ID NO: 3:
34     SEQUENCE CHARACTERISTICS:
35     LENGTH: 4010 base pairs
36     TYPE: nucleic acid
37     STRANDEDNESS: double
38     TOPOLOGY: linear
39     MOLECULE TYPE: cDNA
40     US-08-785-310A-3

```


Best Local Similarity 63.0%; Pred. No. 35;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 2 ATGCTGAGACTAGTATCCGCCGTCATCATGACATCATCTACTGAGAG 55
DB 1829 ATCTGAGAGCATATTCGGTGGCAAGAGAGCTCCCAAGATCCAGAG 1882

RESULT 5

US-09-134-001C-143
Sequence 143, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 143
LENGTH: 1515
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-143

Query Match 33.5%; Score 21.8; DB 4; Length 1515;
Best Local Similarity 70.7%; Pred. No. 33;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2 ATGCTGAGACTAGTATCCGCCGTCATCATGACATCATCTACTGAG 42
DB 497 ATGCTGAGACTATCTATACGCCATCATCAATGTAAATTA 537

RESULT 6

US-08-591-079-9/c
Sequence 9, Application US/08591079
Patent No. 5972899
GENERAL INFORMATION:
APPLICANT: Zychlinksky, Arturo
TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1812
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/591.079
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Llynae, Shmuel
REGISTRATION NUMBER: 33,949
REFERENCE/DOCKET NUMBER: 15661-20017.00
TELEPHONE: (202) 887-1500
TELEFAX: (202) 887-0764
TELEX: 90-4030 MRSNPOERSWSH
INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 5393 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Salmonella typhi
STRAIN: Ty2
FEATURE:
NAME/KEY: CDS
LOCATION: 543..2324
OTHER INFORMATION: /gene= "sipB"
US-08-591-079-9

Query Match 33.5%; Score 21.8; DB 2; Length 5393;
Best Local Similarity 61.4%; Pred. No. 45;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 CATGCTGAGACTAGTATCCGCCGTCATCATGACATCATCAGACTGAGAGAT 57
DB 3822 CTGCTGATTATTTAGTACCGCCCTGCATATGAAATGATATCTGTCCCGATGTT 3766

RESULT 7

PCT-US96-03916-1/c
Sequence 1, Application PC/TUS9603916
GENERAL INFORMATION:
APPLICANT: Wild, Martha A.
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/03916
FILING DATE: 23-MAR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,597
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 39116-A
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 13473 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1059..2489
FEATURE:
NAME/KEY: CDS
LOCATION: 2575..4107

FEATURE:
NAME/KEY: CDS
LOCATION: 4113..4445
FEATURE:
NAME/KEY: CDS
LOCATION: 4609..5487
FEATURE:
NAME/KEY: CDS
LOCATION: 5697..8654
FEATURE:
NAME/KEY: CDS
LOCATION: 9874..10962
FEATURE:
NAME/KEY: CDS
LOCATION: 11159..12658
FEATURE:
NAME/KEY: CDS
LOCATION: 12665..13447
PCT-US96-03916-1

Query Match 33.5%; Score 21.8; DB 5; Length 13473;
Best Local Similarity 65.3%; Pred. No. 56;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 16 TATCCGCGTCATCATGACATCATCACAGTACTGAGGAGATGATAGC 64
Db 11397 TTTCCTCCGCTGTAATTAATCATGAAAGCAATAGTTAGCGAAGGC 11349

RESULT 8
PCT-US96-03916-59/c
Sequence 59, Application PC/TUS9603916

GENERAL INFORMATION:

APPLICANT: Wild, Martha A.

APPLICANT: Cochran, Mark D.

TITLE OF INVENTION: RECOMBINANT INFECTIOUS LARYNGOTRACHEITIS VIRUS

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 72

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patgenin Release #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/03916

FILING DATE: 23-MAR-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/126,597

FILING DATE: 24-SEP-1993

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 39116-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 59:

SEQUENCE CHARACTERISTICS:

LENGTH: 18912 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: linear

HYPOTHEICAL: N

ANTI-SENSE: N

FEATURE:
NAME/KEY: CDS
LOCATION: 697..1533
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1900..2784)
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (2916..3605)
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 3694..5124
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 5210..7081
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 7245..8123
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 8333..11290
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 11098..12402
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 12510..13598
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 13792..15291
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 15298..16080
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: 16129..17013
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS
LOCATION: complement (17380..18216)
OTHER INFORMATION:
PCT-US96-03916-59

Query Match 33.5%; Score 21.8; DB 5; Length 18912;
Best Local Similarity 65.3%; Pred. No. 61;
Matches 32; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 16 TATCCGCGTCATCATGACATCATCACAGTACTGAGGAGATGATAGC 64
Db 14030 TTTCCTCCGCTGTAATTAATCATGAAAGCAATAGTTAGCGAAGGC 13982

RESULT 9
US-08-484-105-1/c
Sequence 1, Application US/08484105

Patent No. 5589341

GENERAL INFORMATION:

APPLICANT: STILLMAN, Bruce

APPLICANT: BELL, Stephen P.

APPLICANT: KOBAYASHI, Ryuji

APPLICANT: RINE, Jasper

APPLICANT: FOSS, Margit

APPLICANT: Abad et al.
TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
FILE REFERENCE: P3393D1
CURRENT APPLICATION NUMBER: US/09/419,291A
PRIOR FILING DATE: 1999-10-15, 450
PRIOR APPLICATION NUMBER: US 09/112,450
PRIOR FILING DATE: 1998-07-09
PRIOR APPLICATION NUMBER: US 60/074,308
PRIOR FILING DATE: 1998-02-11
PRIOR APPLICATION NUMBER: US 60/052,273
PRIOR FILING DATE: 1997-07-10
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patent version 3.1
SEQ ID NO 3
LENGTH: 8561
TYPE: DNA
ORGANISM: Candida albicans
US-09-419-291A-3

Query Match 33.2%; Score 21.6; DB 4; Length 8561;
Best Local Similarity 85.7%; Pred. No. 60;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 25 TCATCATGACATCATCATGACTGAG 52
DB 5413 TAATCATGACACATGACATGAG 5386

RESULT 13
US-09-641-638-111/c
Sequence 111, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET 051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 111
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 501
OTHER INFORMATION: 12-406-52 : polymorphic base C or T
NAME/KEY: misc_binding
LOCATION: 481..500
OTHER INFORMATION: 12-406-52.misl, potential
NAME/KEY: misc_binding
LOCATION: 502..521
OTHER INFORMATION: 12-406-52.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 450..468
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 881..889
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding

LOCATION: 489..513
OTHER INFORMATION: 12-406-52 potential probe
NAME/KEY: misc feature
LOCATION: 4,53,104,243,324,369
OTHER INFORMATION: n=a, g, c or t
US-09-641-638-111

Query Match 32.6%; Score 21.2; DB 4; Length 1001;
Best Local Similarity 69.0%; Pred. No. 50;
Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 24 GTCAATGACATCATCATGAGTGAATGACT 65
DB 845 GTGAAATTAACATCAACAGAGCTGAGAGGATTTCT 804

RESULT 14
US-09-641-638-112/c
Sequence 112, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET 051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 112
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 501
OTHER INFORMATION: 12-406-409 : polymorphic base A or G
NAME/KEY: misc_binding
LOCATION: 481..500
OTHER INFORMATION: 12-406-409.misl, potential
NAME/KEY: misc_binding
LOCATION: 502..521
OTHER INFORMATION: 12-406-409.mis2, potential complement
NAME/KEY: primer_bind
LOCATION: 93..111
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 524..542
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 489..513
OTHER INFORMATION: 12-406-409 potential probe
NAME/KEY: misc feature
LOCATION: 12,717,742,755,758,1000..1001
OTHER INFORMATION: n=a, g, c or t
US-09-641-638-112

Query Match 32.6%; Score 21.2; DB 4; Length 1001;
Best Local Similarity 69.0%; Pred. No. 50;
Matches 29; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 24 GTCAATGACATCATCATGAGTGAATGACT 65

Db 488 GTGAAATATACATCAAAACAGACTGTGAGAGAGAAATTCT 447

RESULT 15

US-09-362-831-10/c
; Sequence 10, Application US/09362831
; Patent No. 6306400
; GENERAL INFORMATION:
; APPLICANT: BUBLOT et al.
; TITLE OF INVENTION: AVIAN RECOMBINANT LIVE VACCINE USING, AS VECTOR, THE
; TITLE OF INVENTION: AVIAN INFECTIOUS LARYNGOTRACHEITIS VIRUS
; FILE REFERENCE: 454313-2520
; CURRENT APPLICATION NUMBER: US/09/362,831
; CURRENT FILING DATE: 1999-07-28
; NUMBER OF SEQ. ID NOS: 28
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 10
; LENGTH: 3116
; TYPE: DNA
; ORGANISM: Infectious laryngotracheitis Virus
US-09-362-831-10

Query Match 32.6%; Score 21.2; DB 4; Length 3116;
Best Local Similarity 64.0%; Pred. No. 66;
Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 2 ATGCTGAGACTATATCCGCCGATCAATGACATCATCAGTACTGA 51
Db 2153 ATGCTGAGAGAGATTCCTTATCATCATGATATCATCAATCATCA 2104

Search completed: May 11, 2003, 03:07:06
Job time : 38.0652 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 1044.89 Seconds
(Without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65
Sequence: 1 catgctgcagactagatcc.....tactgagagatgatagct 65

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: em_estda:*
2: em_esthm:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	28.4	43.7	614	12	BF168473 601776065
C 2	27.4	42.2	563	17	BH372871 AG-ND-106
C 3	27.4	42.2	665	17	BH365471 AG-ND-106
C 4	27.4	42.2	814	17	BH371422 AG-ND-172
C 5	27	41.5	785	13	B1971380 GM830013A
C 6	26.8	41.2	769	12	BG433137 602496882

C 7	26.8	41.2	968	11	BC034140	BC034140 Homo sapi
C 8	26.4	40.6	543	12	BE975786	BE975786 b545c02.x
C 9	26.2	40.3	772	17	CNS015AK	AL164981 Tetradodon
C 10	26	40.0	103	12	BF466720	BF466720 UI-M-CGDP
C 11	26	40.0	135	14	BM942017	BM942017 UI-M-CGDP
C 12	26	40.0	103	14	BE531686	BE531686 601230802
C 13	26	40.0	549	14	BQ562476	BQ562476 H4076B12-
C 14	26	40.0	549	13	BI713338	BI713328 IC85H09.Y
C 15	26	40.0	566	13	BI715981	BI715981 IC64D03.Y
C 16	26	40.0	613	10	BB621994	BB621994 BB621994
C 17	26	40.0	617	10	BB659983	BB659983 BB659983
C 18	26	40.0	618	10	BB618417	BB618417 BB618417
C 19	26	40.0	618	10	BB651874	BB651874 BB651874
C 20	26	40.0	642	10	BB655827	BB655827 BB655827
C 21	26	40.0	654	10	BB638327	BB638327 BB638327
C 22	26	40.0	663	12	BF348858	BF348858 RCL-PT002
C 23	26	40.0	703	9	AT1747024	AT1747024 U112604.Y
C 24	26	40.0	748	13	BG969043	BG969043 602834988
C 25	26	40.0	768	14	BQ745618	BQ745618 UI-M-EMO-
C 26	26	40.0	868	13	BI739607	BI739607 603361862
C 27	26	40.0	938	12	BG293515	BG293515 602390444
C 28	26	40.0	954	14	BQ884980	BQ884980 AGENCOURT
C 29	26	40.0	973	14	BQ898008	BQ898008 AGENCOURT
C 30	25.8	39.7	388	9	AI344310	AI344310 TC03d11.x
C 31	25.8	39.7	388	9	AI344331	AI344331 TC03d11.x
C 32	25.8	39.7	656	17	AG064170	AG064170 Pan trogl
C 33	25.6	39.4	515	13	BM052014	BM052014 Tc ad2.02
C 34	25.6	39.4	547	14	BQ097606	BQ097606 ph05a07.Y
C 35	25.6	39.4	930	17	CNS07CK4	AL439742 T7 end of
C 36	25.4	39.1	308	17	AZ218097	AZ218097 Sheared D
C 37	25.4	39.1	524	17	AZ217624	AZ217624 Sheared D
C 38	25.2	38.8	767	12	BE748221	BE748221 601571441
C 39	25	38.5	379	9	AA235966	AA235966 z805f02.8
C 40	25	38.5	381	9	AA814930	AA814930 oc07c03.s
C 41	25	38.5	438	10	AW971534	AW971534 RST383623
C 42	24.8	38.2	495	14	BQ378960	BQ378960 RC3-UT006
C 43	24.8	38.2	504	9	AA542644	AA542644 fa08e10.r
C 44	24.8	38.2	530	9	AA390522	AA390522 LD08708.5
C 45	24.8	38.2	619	17	AZ952780	AZ952780 2M0217C18

ALIGNMENTS

RESULT 1
LOCUS BF168473/c 614 bp mRNA linear EST 30-OCT-2000
DEFINITION 601776065F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:4017668 5',
mRNA sequence.
ACCESSION BF168473
VERSION BF168473.1 GI:11048825
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 614)
NHI-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Straubeberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LHAM267 row: a column: 17
High quality sequence stop: 614.
Location/Qualifiers
1. 614

FEATURES
source

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:4017688"
 /clone_lib="NCI CGAP Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary."
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI. Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 198 a 97 c 154 g 164 t 1 others
 ORIGIN

Query Match 43.7%; Score 28.4; DB 12; Length 614;
 Best Local Similarity 65.1%; Pred. No. 18;
 Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 3 TGCTGACAGACTAGTATCCGCGCTCATCATGATCATCATGAGTACTGAGAGATGAATA 62
 DB 173 TTCTGCACTCTGTTAAAGCCGATCTTCTATATCAACTTAACAGACAGAGAAATGACCA 114

QY 63 GC 64
 DB 113 GC 112

RESULT 2
 BH372871/c 563 bp DNA linear GSS 10-DEC-2001
 LOCUS AG-ND-106G18.TR ND-TAM Anopheles gambiae genomic clone AG-ND-106G18
 DEFINITION , DNA sequence.
 ACCESSION BH372871 GI:17319013
 VERSION BH372871
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.

REFERENCE 1 (bases 1 to 563)
 AUTHORS Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: AG-ND-106G18.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..563

/organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-106G18"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site 1: HindIII"
 BASE COUNT 157 a 107 c 83 g 216 t

ORIGIN

Query Match 42.2%; Score 27.4; DB 17; Length 563;
 Best Local Similarity 65.6%; Pred. No. 38;
 Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 2 ATGCTGACAGACTAGTATCCGCGCTCATCATGATCATCATGAGTACTGAGAGATGAAT 61
 DB 342 ATTTGACAGAGAGTAAGACCTTAGAGAGAGACAAATATGATTAAGAGGAATATAT 283

QY 62 A 62
 DB 282 A 282

RESULT 3
 BH385471/c 665 bp DNA linear GSS 10-DEC-2001
 LOCUS AG-ND-106B1.TR ND-TAM Anopheles gambiae genomic clone AG-ND-106B1.
 DEFINITION , DNA sequence.
 ACCESSION BH385471 GI:17331613
 VERSION BH385471
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.

REFERENCE 1 (bases 1 to 665)
 AUTHORS Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: AG-ND-106B1.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..665

/organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-106B1"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site 1: HindIII"
 BASE COUNT 196 a 119 c 100 g 250 t

Query Match 42.2%; Score 27.4; DB 17; Length 665;
 Best Local Similarity 65.6%; Pred. No. 41;
 Matches 40; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 2 ATGCTGACAGACTAGTATCCGCGCTCATCATGATCATCATGAGTACTGAGAGATGAAT 61
 DB 341 ATTTGACAGAGAGTAAGACCTTAGAGAGAGACAAATATGATTAAGAGGAATATAT 282

QY 62 A 62
 DB 281 A 281

RESULT 4	BH371422/c	BH371422	814 bp	DNA	linear	GSS 10-DEC-2001
LOCUS	AG-ND-172C15.TF	ND-TAM	Anopheles gambiae	genomic clone	AG-ND-172C15	
DEFINITION	, DNA sequence.					
ACCESSION	BH371422					
VERSION	BH371422.1	GI:17317547				
KEYWORDS	GSS.					
SOURCE	African malaria mosquito.					
ORGANISM	Anopheles gambiae					
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.					
AUTHORS	1 (bases 1 to 814)					
TITLE	Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.					
JOURNAL	Direct Submission of BAC-end sequences from Anopheles gambiae unpublished (2001)					
COMMENT	Other GSSs: AG-ND-172C15.TF					
	Contact: Brendan J Loftus					
	Department of Eukaryotic Genomics					
	The Institute for Genomic Research					
	9712 Medical Center Dr., Rockville, MD 20850, USA					
	Tel: 301 838 0208					
	Fax: 301 838 3543					
	Email: b.loftus@tigr.org					
	This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.					
	Seq primer: M13 For					
	Class: BAC ends.					
FEATURES	Location/Qualifiers					
source	1..814					
	/organism="Anopheles gambiae"					
	/strain="PEST"					
	/db_xref="taxon:7165"					
	/clone="AG-ND-172C15"					
	/clone_11b="ND-TAM"					
	/note="Vector: pECBAC1; Site_1: HindIII"					
BASE COUNT	237 a	141 c	132 g	304 t		
ORIGIN						
Query Match	42.2%	Score 27.4;	DB 17;	Length 814;		
Best Local Similarity	65.6%	Pred. No. 45;				
Matches 40;	Conservative 0;	Mismatches 21;	Indels 0;	Gaps 0;		
CY	2	ATGCGACACATGATCGCGCGTATCATATGATCATCATGATCTAGAGAGATGAT	61			
DB	347	ATTTCGACGACAGTAAAGACCTTACGAGAGACATATTCAGATTTCAGAGAAATGAT	288			
CY	62	A	62			
DB	287	A	287			
RESULT 5	B1971380	785 bp	mRNA	linear	EST 23-OCT-2001	
LOCUS	GMS30013A20F02	Gm-r1083	Glycine max	CDNA clone	Gm-r1083-4852.3	
DEFINITION	mRNA sequence.					
ACCESSION	B1971380					
VERSION	B1971380.1	GI:16345785				
KEYWORDS	EST.					
SOURCE	soybean.					
ORGANISM	Glycine max					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Spermatophytes: Magnoliophyta: eudicotyledons: core eudicots:
Rosidae, eutroside I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 785)
Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V.,
Ergleding, J., Rapp, C., Shoop, E., Padias, J., Liu, L., and Lewin, H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Other ESTs: BE022118 corresponding to Gm-cl028-9003 (5')
Contact: Vodkin, L.O., Pl. A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Beck Center for Comparative and Functional
Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Incyte Genomics, 4633 World
Parkway Circle St. Louis, Missouri 63134. Phone (800) 430-0030 or
(314) 427-3222 FAX: (314) 427-3324. Web site:
<http://www.incyte.com/reagents/catalog.jsp?page=clones/collaboratio>
n/index
Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
Location/Qualifiers
1..785
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="Gm-r1083-4852"
/clone_lib="Gm-r1083"
/note="The library Gm-r1083 is a sequence-driven, rereckered
set of 4,992 clones selected from cDNA libraries from
various tissues and stages of development of soybean. It
represents 1117 sequences from the progenitor library
Gm-cl009 (from mature roots of 2 month old greenhouse
grown 'Williams' soybean plants); 820 sequences from the
progenitor library Gm-cl013 (from 2 to 3 week old whole
plants of Williams) and 3055 sequences from library
Gm-cl028 (from 'Superpod' plants whose seedlings were
immobilized with Bradyrhizobium japonicum, courtesy of Dr.
Gary Stacey). The 5' ESTs of the source clones from the
different progenitor libraries was used to select
singletons, or a representative of each contig, which were
rereckered to form library Gm-r1083. The cDNA clones of the
rereckered Gm-r1083 library were then sequenced at the 3'
end. The contig analysis to select unique genes was
performed by the laboratory of Ernest Retzel, Center for
Computational Genomics and Bioinformatics, University of
Minnesota, <http://web.ahc.umn.edu/biodata/nefsoy/>.
Rerecking was performed by Incyte Genomics, St. Louis,
<http://www.incyte.com>, and 3' sequencing by the Beck
Center for Comparative and Functional Genomics, University
of Illinois, <http://www.lle.uiuc.edu/biotech/beck.html>.
Note: The corresponding 5' EST from each clone in the
Gm-r1083 library is listed in the 'OTHER EST' field. The
detailed information on the source library for each clone
can also be obtained by referring to the Incyte Genomics
clone ID of the original cDNA library that is also listed
under 'OTHER EST'."

BASE COUNT 220 a 177 c 134 g 247 t 7 others
ORIGIN

Query Match 41.5%; Score 27; DB 13; Length 785;
Best Local Similarity 66.1%; Pred. No. 61;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

3 TGTCTGACAGTAAAGTATCCGCGGTATCATATGATCATCATCAAGTACTGAGAGATGAT 61
DB 269 TGTCTGACAGTAAAGTATCCGCGGTATCATATGATCATCATCAAGTACTGAGATGAT 327

LOCUS	BG433137	769 bp	mRNA	linear	EST 14-MAR-2002
DEFINITION	602496882P1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4610636 5', mRNA sequence.				
ACCESSION	BG433137				
VERSION	BG433137.1	GI:13339643			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 769)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgraeb@remail.nih.gov Tissue Procurement: CLONTECH Laboratories, Inc. cDNA Library Preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: http://image.llnl.gov Plate: LHC1356 row: c column: 21 High quality sequence stop: 741. Location/Qualifiers				
FEATURES	source				
	1..769				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:4610636"				
	/clone_lib="NIH_MGC_75"				
	/lab_host="DH10B (T1 phage-resistant)"				
	/note="Organ: kidney; Vector: pDNR-L3B (Clontech); Site:1; SfiI (ggcgccctcgcc); Site:2: SfiI (ggccatctagcc); 5' adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCCCGAGCGCCGACACTG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH MGC library."				
BASE COUNT	215 a 122 c 178 g 253 t 1 others				
ORIGIN					
	Query Match				
	Best Local Similarity 64.5%; Pred. No. 70;				
	Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;				
CY	1	CATGCTGCAGCATAGATCCGCCGTCATTAAGCATCTACAGTACGAGAGATGA	60		
DB	390	CATTACACGAACGATATCAGCACACGAGCATCTGATTAAGTATTGAATTAATGA	331		
CY	61	TA 62			
DB	330	TA 329			
RESULT 7					
LOCUS	BC034140/C	968 bp	mRNA	linear	HTC 08-JUL-2002
DEFINITION	Homo sapiens, similar to likely ortholog of yeast ARV1, clone IMAGE:4610636, mRNA.				
ACCESSION	BC034140				
VERSION	BC034140.1	GI:21707881			
KEYWORDS	HTC.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 968)				
TITLE	Strausberg, R.				
	Direct Submission				

JOURNAL

Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC) Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2530, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

REMARK COMMENT
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH
CLONTECH Laboratories, Inc.
CLONTECH Library Preparation: CLONTECH Laboratories, Inc.
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LILN)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www.sbgc.stanford.edu>
Contact: (Dickson, Mark) mcd@pxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILN at: <http://image.llnl.gov>
Series: IRAL Plate: 41 Row: f Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gl: 7662105
This clone has the following problem: frame shifted.

FEATURES
Source
Location/Qualifiers
1..968
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4610636"
/tissue_type="Kidney"
/clone_lib="NIH-SBGC_75"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"

BASE COUNT 358 a 135 c 196 g 279 t
ORIGIN

Query Match 41.2% Score 26.8; DB 11; Length 968;
Best Local Similarity 64.5%; Pred. No. 78;
Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 1 CATGCTGCAGACTGTATCCCGCCTCATTCATTCATCATCACACTGACGAGATCAA 60
DB 389 CATTCAACAAGAAGTAGTATCAGACACACAGCAGCAGCACTTGATTAAGTAAATGAATCA 330

OY 61 TA 62
DB 329 TA 328

RESULT 8
BE975786/c 543 bp mRNA linear EST 04-OCT-2000
LOCUS bsa5c02.x1 Drosophila melanogaster adult testis library Drosophila
DEFINITION melanogaster cDNA clone bsa5c02.3', mRNA sequence.
ACCESSION BE975786
VERSION BE975786.1 GI:10604624
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 543)
Andrews, J., Bouffard, G. and Oliver, B.
Drosophila melanogaster testis expressed sequence tags
Unpublished (1999)
Contact: Brian Oliver
Laboratory of Cellular and Developmental Biology
NIDDK, National Institutes of Health
6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
Fax: (301) 496 5239
Email: oliver@helix.nih.gov,
<http://www.niddk.nih.gov/intram/people/bolliver.htm>
Tissue isolation and library construction performed at the National

Institute of Diabetes and Digestive and Kidney Diseases, NIH (see <http://www.nidck.nih.gov/intram/people/boliver.htm>). DNA sequencing and analyses performed by National Institutes of Health Intramural Sequencing Center (NISC, see <http://www.nisc.nih.gov>).

Plate: 45 row: C column: 02
Seq primer: -21M13 forward primer (ABI).

Location/Qualifiers

source

1. 543

/organism="Drosophila melanogaster"

/strain="y[+] w[67c1]/Y"

/db_xref="taxon:7227"

/clone_1ib="Ds45C02"

/clone_1ib="Drosophila melanogaster adult testis library"

/sex="male"

/dev_stage="1-5 day adult"

/lab_host="SOLR (Stratagene)"

/note="Organ: testis; Vector: pBluescript SK (Stratagene); Site: 1: Bcor I; Site 2: Xho I; Testes dissected from 1-5 day adult y[+] w[67c1]/Y males raised at 25°C. RNA isolated using Trizol (Life Technologies) and a single round of Poly(A)+ selection using Oligotex (Qiagen). cDNA library constructed using Stratagene ZAP-cDNA synthesis kit. Oligo dt-primed, size fractionated -1-6 kb, and directionally cloned at EcoRI and XhoI in Uni-ZAP XR. Following a single round of amplification pBluescript SK phagemids were mass excised. A distribution channel for clones is being sought, but not currently available. Requests for clones cannot be honored."

BASE COUNT 139 a 111 c 85 g 208 t

ORIGIN

Query Match

Best Local Similarity 40.6%; Score 26.4; DB 12; Length 543; Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 CATGCTGACAGTACTGTCGCGCCGTCATCATGATCATGATGAGATGAA 60
DB 160 CATGAGTGTGACATGAGTCCGACGATCATGATGATTAATCTAGAAAAAGAGCGAGAA 101

RESULT 9

CNS01SAK

LOCUS

DEFINITION

CNS01SAK 772 bp DNA linear GSS 12-MAY-2000
Tetradon nigroviridis genome survey sequence PUC-Or1 end of clone 169B05 of library G from Tetradon nigroviridis, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetradon>.

FEATURES

source

1. 772

/organism="Tetradon nigroviridis"

/db_xref="taxon:9983"

/clone_1ib="G"

/clone_1ib="G"

/note="Genoscope sequence ID : C0AG169CA03SP1-end : PUC-Or1"

BASE COUNT

ORIGIN

Query Match 40.3%; Score 26.2; DB 17; Length 772; Best Local Similarity 64.9%; Pred. No. 1.1e+02; Matches 37; Conservative 1; Mismatches 19; Indels 0; Gaps 0;

QY 5 CTGACAGTGTATGCGCCGTCATCATGATCATGATGAGATGAA 61
DB 289 CTGACAGTGTATGCGCCGTCATCATGATCATGATGAGATGAA 345

RESULT 10

BP466720

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Genome Res. 6 (9), 791-806 (1996)

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: MEST@mail.nih.gov

Oligo-dT track not found. Not 1 site shown in beginning of sequence

is likely internal to the message. cDNA library preparation: M.B.

Soares Lab Clone distribution: Researchers may obtain BMAP cDNA

clones from RESEARCH GENETICS. It should be noted that Bento Soares

is generating a small number of additional specialized

non-redundant arrays of BMAP cDNAs whose availability will be

considered under appropriate and limited collaborative arrangements

Seq primer: M13 Forward

FOLTA=NO.

Location/Qualifiers

1. 103

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_1ib="NIH BMAP Ret4_S2"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; The

NIH BMAP Ret4_S2 library is a subtracted library,

ultimately derived from mouse retina tissue libraries at

various stages of development. For a detailed description

of the library from which this clone was derived, please

visit our web site at brainest.eng.uiowa.edu. The tissue

for this library was contributed by Dr. Xin-Yuan Fu, Yale University School of Medicine
TAG_S80=None found"

BASE COUNT 27 a 23 c 30 g 23 t

Query Match 40.0%; Score 26; DB 12; Length 103;
Best Local Similarity 70.0%; Pred. No. 54;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 4 GCTGCAGACTAGTATCCGCGCATCATGACATCATGACTGAGG 53
DB 27 GCTGCAGAAATAGGAAAGCCGCGCATGATGAGTCCACAGACTATGTGG 76

RESULT 11
BM942017/c 135 bp mRNA linear EST 29-APR-2002

LOCUS
DEFINITION
UI-M-CGDP-bqx-f-04-0-UI.r1 NIH BMAP Ret4 S2 Mus musculus cDNA clone
BM942017
BM942017.1 GI:19401397

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 135)
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestremail.nih.gov

Tissue Procurement: Dr. Xin-Yuan Fu, Yale University School of
Medicine

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.regen.com).

Seq primer: M13 REVERSE.

Location/Qualifiers

1. .135
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-CGDP-bqx-f-04-0-UI"
/clone_1b="NIH BMAP Ret4 S2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The
NIH BMAP Ret4 S2 library is a subtracted library,
ultimately derived from mouse retina tissue libraries at
various stages of development. For a detailed description
of the library from which this clone was derived, please
visit our web site at brainest.eng.uiowa.edu. The tissue
for this library was contributed by Dr. Xin-Yuan Fu, Yale
University School of Medicine"

BASE COUNT 35 a 29 c 32 g 32 t

Query Match 40.0%; Score 26; DB 14; Length 135;

Best Local Similarity 70.0%; Pred. No. 61;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 4 GCTGCAGACTAGTATCCGCGCATCATGACATCATGACTGAGG 53

DB 68 GCTGCAGAAATAGGAAAGCCGCGCATGATGAGTCCACAGACTATGTGG 19

RESULT 12
BE531686/c 482 bp mRNA linear EST 09-AUG-2000
LOCUS
DEFINITION
601230802F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3594577 5',
mRNA sequence.
BE531686
BE531686.1 GI:9760331

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 482)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L14M8769 row: d column: 02
High quality sequence stop: 482.

FEATURES

source
1. .482
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3594577"
/clone_1b="NCI CGAP Mam6"
/sex="Female, virgin"
/tissue_type="Infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 140 a 113 c 97 g 132 t

Query Match 40.0%; Score 26; DB 10; Length 482;
Best Local Similarity 70.0%; Pred. No. 11e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 4 GCTGCAGACTAGTATCCGCGCATCATGACATCATGACTGAGG 53
DB 270 GCTGCAGAAATAGGAAAGCCGCGCATGATGAGTCCACAGACTATGTGG 221

RESULT 13
B0562476/c 545 bp mRNA linear EST 20-JUN-2002

LOCUS
DEFINITION
H4076B12-5 NIA Mouse 7.4K cDNA clone Set Mus musculus cDNA clone
H4076B12 5', mRNA sequence.
B0562476
B0562476.1 GI:21463362

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 545)
VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G., Martin
P.R., Stags,C.A., Bassey,U., Alba,K., Hamatani,T., Kargul,G.J.,

TITLE Luo, A.G. and Ko, M.S.H.
Assembly, verification, and initial annotation of NIA 7.4K mouse
cDNA clone set
JOURNAL Unpublished (2002)
COMMENT Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@imgn.gsc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://imgn.gsc.nia.nih.gov/cDNA/NIA_7_4k.html for details.
Plate: H4076 Row: B Column: 12
Seq primer: -21M13 Reverse
High quality sequence stop: 545
POLYA=No.

FEATURES

source Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/db_xref="IMAGE:5660776"
/clone_lib="NIA Mouse 7.4K cDNA Clone Set"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This
clone is among a rearranged set of 7,407 clones from more
than 20 cDNA libraries."
BASE COUNT 176 a 113 c 104 g 152 t
ORIGIN

Query Match 40.0%; Score 26; DB 14; Length 545;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 4 GCTGCAGACTAGTATCCGCGTCATCATGACATCATCAGACTAGTGG 53
160 GCTGCAGACTAGTATCCGCGTCATCATGACATCATCAGACTAGTGG 111

RESULT 14 549 bp mRNA linear EST 12-MAR-2002
LOCUS B1713238/c
DEFINITION Musculus cDNA clone IMAGE:5660776 5' similar to TR:Q9VZ81 Q9VZ81
CG13708 PROTEIN.; mRNA sequence.
ACCESSION B1713238 GI:15688933
VERSION B1713238
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 549)

REFERENCE Melton, D., Brown, J., Keny, G., Permutt, A., Lee, C., Kaestner, K.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistein, A.,
Schmitt, A., Theising, B., Ritzer, E., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T.,
Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Other ESTs: ic85h09.xl
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@iobp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:

Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brownjefas.harvard.edu)
MGI:1947102 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 429.
Location/Qualifiers

FEATURES

source 1..549
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
/clone_lib="IMAGE:5660776"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
NI-MMSI"
/sex="Both for embryonic & newborn, male for adult and
adult islet."
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; Five
libraries representing E10.5/12.5 pancreatic bud, E16.5
pancreas, newborn pancreas, adult pancreas, and adult
islets of Langerhans were separately constructed using
Superscript Plasmid Library kit (Life Technologies). cDNA
was made by oligo-dT priming and size-selected by column
fractionation. Libraries were amplified once on solid
support and plasmid DNA from each library was prepared
and mixed in equal amounts. The mixed library DNA was
normalized by method #4 from Bonaldo, Lennon, and Soares
1996 Genome Research 6:791-806; 0.5 microgram
single-stranded mixed library plasmid DNA was mixed with
5 micrograms PCR product representing mixed library
inserts and hybridized to an EcoT of 6. Single-stranded
(unhybridized) plasmids were isolated by hydroxyapatite
chromatography and used to make this library."
BASE COUNT 163 a 131 c 110 g 145 t
ORIGIN

Query Match 40.0%; Score 26; DB 13; Length 549;
Best Local Similarity 70.0%; Pred. No. 1.1e+02;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
QY 4 GCTGCAGACTAGTATCCGCGTCATCATGACATCATCAGACTAGTGG 53
256 GCTGCAGACTAGTATCCGCGTCATCATGACATCATCAGACTAGTGG 207

RESULT 15 566 bp mRNA linear EST 12-MAR-2002
LOCUS B1715981/c
DEFINITION Musculus cDNA clone IMAGE:5658197 5' similar to TR:Q9VZ81 Q9VZ81
CG13708 PROTEIN.; mRNA sequence.
ACCESSION B1715981 GI:15691676
VERSION B1715981
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 566)

REFERENCE Melton, D., Brown, J., Keny, G., Permutt, A., Lee, C., Kaestner, K.,
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistein, A.,
Schmitt, A., Theising, B., Ritzer, E., Ronko, I., Bennett, J., Cardenas
, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T.,
Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute

TITLE
JOURNAL
COMMENT

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu)

MG1:1944523 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov

Seq primer: -408P from Gibco

High quality sequence atop: 428.

Location/Qualifiers

1. 566

/organism="Mus musculus"

/strain="ICR"

/db_xref="taxon:10090"

/clone="IMAGE:5658197"

/clone_lib="Melton Normalized Mixed Mouse Pancreas 1

N1-WMS1"

/sex="Both for embryonic & newborn, male for adult and

adult islet"

/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,

adult, mixed"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site 1: Not I; Site 2: Sal I; Five

libraries representing E10.5/12.5 pancreatic bud, E16.5

pancreas, newborn pancreas, adult pancreas, and adult

islets of Langerhans were separately constructed using

SuperScript Plasmid Library kit (Life Technologies). cDNA

was made by oligo-dT priming and size-selected by column

fractionation. Libraries were amplified once on solid

support and plasmid DNA from each library was prepared

and mixed in equal amounts. The mixed library DNA was

normalized by method #4 from Bonaldo, Lennon, and Soares

1996 Genome Research 6:791-806; 0.5 microgram

single-stranded mixed library plasmid DNA was mixed with

5 micrograms PCR product representing mixed library

inserts and hybridized to an EcoT of 6. Single-stranded

(unhybridized) plasmids were isolated by hydroxyapatite

chromatography and used to make this library."

BASE COUNT 180 a 121 c 113 g 152 t

ORIGIN

Query Match 40.0%; Score 26; DB 13; Length 566;

Best Local Similarity 70.0%; Freq. No. 1.2e+02;

Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 4 GCTGCAGACTAGTATCCGCGTCATCATCATCATCACTACTGAGG 53

DB 228 GCTGCAGATAGGAACCGCGCATGATGACTCACCAGACTATGTGG 179

Search completed: May 11, 2003, 04:54:08
Job time : 1049.89 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:08:16 ; Search time 64.3484 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-8

Perfect score: 65
Sequence: 1 catgctgcagactagatcc.....tactgagagatgatagct 65

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	100.0	65	9	US-09-963-803-8
2	65	100.0	317	9	US-09-963-803-3
3	65	100.0	371	9	US-09-963-803-5
4	58	89.2	348	9	US-09-963-803-4
5	58	89.2	398	9	US-09-963-803-6
6	58	89.2	472	9	US-09-963-803-25
7	58	89.2	541	9	US-09-963-803-24
8	58	89.2	604	9	US-09-963-803-21
9	55	84.6	392	9	US-09-963-803-21
10	55	84.6	393	9	US-09-963-803-19
11	55	84.6	462	9	US-09-963-803-20
12	55	84.6	600	9	US-09-963-803-22
13	46	70.8	243	9	US-09-963-803-1
14	24.2	37.2	2600	10	US-09-954-456-1157
15	23.6	36.3	273	9	US-10-015-219-232
16	23.6	36.3	273	10	US-09-777-564-232
17	22.6	34.8	459	10	US-09-864-761-3604
18	22.4	34.5	438	9	US-09-918-995-4240
19	22.4	34.5	906	9	US-09-938-842A-2162

20	22.4	34.5	1170	10	US-09-974-300-1705	Sequence 1705, Appl
21	22.4	34.5	1902	9	US-09-938-842A-521	Sequence 521, Appl
22	22.4	34.5	3393	9	US-09-738-626-2262	Sequence 2262, Ap
23	22.4	34.5	3309400	9	US-09-738-626-1	Sequence 1, Appli
24	22.2	34.2	1086	9	US-09-991-496-77	Sequence 77, Appli
25	22.2	34.2	1086	10	US-09-874-923-77	Sequence 77, Applh
26	22.2	34.2	1179	10	US-09-759-097-1	Sequence 1, Appli
27	22.2	34.2	1533	12	US-10-074-547-3	Sequence 3, Appli
28	22.2	34.2	4419	12	US-10-074-547-1	Sequence 1, Appli
29	22	33.8	42000	9	US-10-081-563-25	Sequence 25, Appl
30	22	33.8	402850	9	US-09-844-553-5	Sequence 5, Appli
31	22	33.8	465237	10	US-09-933-267A-1	Sequence 1, Appli
32	21.8	33.5	218	10	US-09-923-876-2430	Sequence 2430, Ap
33	21.8	33.5	1740	9	US-09-922-364A-15	Sequence 15, Appl
34	21.8	33.5	1740	9	US-09-254-590-15	Sequence 15, Appl
35	21.8	33.5	1740	9	US-10-115-695-15	Sequence 15, Appl
36	21.8	33.5	1740	9	US-10-116-561-15	Sequence 15, Appl
37	21.8	33.5	1740	9	US-10-115-671-15	Sequence 15, Appl
38	21.8	33.5	1740	9	US-10-115-415-15	Sequence 15, Appl
39	21.8	33.5	1740	9	US-10-116-260-15	Sequence 15, Appl
40	21.8	33.5	3605	10	US-09-881-457A-3	Sequence 3, Appli
41	21.8	33.5	13473	9	US-09-994-064-1	Sequence 1, Appli
42	21.8	33.5	18913	9	US-09-994-064-59	Sequence 59, Appl
43	21.6	33.2	279	10	US-09-864-761-21997	Sequence 21997, Ap
44	21.6	33.2	315	9	US-09-938-842A-1593	Sequence 1593, Ap
45	21.6	33.2	473	10	US-09-864-761-6211	Sequence 6211, Ap

ALIGNMENTS

RESULT 1
US-09-963-803-8
Sequence 8, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963.803
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
LENGTH: 65
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: S1
US-09-963-803-8

Query Match 100.0%; Score 65; DB 9; Length 65;
Best Local Similarity 100.0%; Pred. No. 6.4e-16;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGTCGACGACTGATCCGCCGCATCATGACATCATGACTAGTACGAGATGAA 60
DB 1 CATGTCGACGACTGATCCGCCGCATCATGACATCATGACTAGTACGAGATGAA 60

QY 61 TAGCT 65
DB 61 TAGCT 65

RESULT 2
US-09-963-803-3
Sequence 3, Application US/09963803

Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 317
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1116
NAME/KEY: Promoter
LOCATION: (1)..(317)
OTHER INFORMATION:
US-09-963-803-3

Query Match 100.0%; Score 65; DB 9; Length 317;
Best Local Similarity 100.0%; Pred. No. 1e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATGCTGAGACTAGTATCCGCCGTCAATGACATCATGACATGACAGTACTGAGAGATGAA 60
DB 8 CATGCTGAGACTAGTATCCGCCGTCAATGACATCATGACATCATGACAGTACTGAGAGATGAA 67

QY 61 TAGCT 65
DB 68 TAGCT 72

RESULT 3
US-09-963-803-5
Sequence 5; Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 371
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1146
NAME/KEY: Promoter
LOCATION: (1)..(371)
OTHER INFORMATION:
US-09-963-803-5

Query Match 100.0%; Score 65; DB 9; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.1e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGCTGAGACTAGTATCCGCCGTCAATGACATCATGACATGACAGTACTGAGAGATGAA 60

DB 8 CATGCTGAGACTAGTATCCGCCGTCAATGACATCATGACATGACAGTACTGAGAGATGAA 67
QY 61 TAGCT 65
DB 68 TAGCT 72

RESULT 4
US-09-963-803-4
Sequence 4; Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 348
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1117
NAME/KEY: Promoter
LOCATION: (1)..(348)
OTHER INFORMATION:
US-09-963-803-4

Query Match 89.2%; Score 58; DB 9; Length 348;
Best Local Similarity 100.0%; Pred. No. 5.7e-13;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCCGTCAATGACATCATGACATGACAGTACTGAGAGATGAGT 65
DB 49 CAGACTAGTATCCGCCGTCAATGACATCATGACATGACAGTACTGAGAGATGAGT 106

RESULT 5
US-09-963-803-6
Sequence 6; Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US/09/963,803
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 398
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1147
NAME/KEY: Promoter
LOCATION: (1)..(398)
OTHER INFORMATION:

US-09-963-803-6

Query Match 89.2%; Score 58; DB 9; Length 398;

Best Local Similarity 100.0%; Pred. No. 5.9e-13; Indels 0; Gaps 0;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 65
|||||
DB 49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 106

RESULT 6

US-09-963-803-25

Sequence 25, Application US/09963803

Publication No. US20030028922A1

GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics

TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow

TITLE OF INVENTION: virus and cassava vein mosaic virus

FILE REFERENCE: 184332042

CURRENT APPLICATION NUMBER: US/09/963,803

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29

PRIOR APPLICATION NUMBER: PCT IB00/00370

PRIOR FILING DATE: 2000-10-05

NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn version 3.1

SEQ ID NO 25

LENGTH: 472

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: promoter MPr1169

NAME/KEY: promoter

LOCATION: (1)..(472)

OTHER INFORMATION:

US-09-963-803-25

Query Match 89.2%; Score 58; DB 9; Length 472;

Best Local Similarity 100.0%; Pred. No. 6.2e-13;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 65
|||||
DB 49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 106

RESULT 7

US-09-963-803-24

Sequence 24, Application US/09963803

Publication No. US20030028922A1

GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics

TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow

TITLE OF INVENTION: virus and cassava vein mosaic virus

FILE REFERENCE: 184332042

CURRENT APPLICATION NUMBER: US/09/963,803

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29

PRIOR APPLICATION NUMBER: PCT IB00/00370

PRIOR FILING DATE: 2000-10-05

NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn version 3.1

SEQ ID NO 24

LENGTH: 541

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: promoter MPr1168

NAME/KEY: promoter
LOCATION: (1)..(541)
OTHER INFORMATION:

US-09-963-803-24

Query Match 89.2%; Score 58; DB 9; Length 541;

Best Local Similarity 100.0%; Pred. No. 6.5e-13;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 65
|||||
DB 49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 106

RESULT 8

US-09-963-803-23

Sequence 23, Application US/09963803

Publication No. US20030028922A1

GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics

TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow

TITLE OF INVENTION: virus and cassava vein mosaic virus

FILE REFERENCE: 184332042

CURRENT APPLICATION NUMBER: US/09/963,803

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29

PRIOR APPLICATION NUMBER: PCT IB00/00370

PRIOR FILING DATE: 2000-10-05

NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn version 3.1

SEQ ID NO 23

LENGTH: 604

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: promoter MPr1167

NAME/KEY: promoter

LOCATION: (1)..(604)

OTHER INFORMATION:

US-09-963-803-23

Query Match 89.2%; Score 58; DB 9; Length 604;

Best Local Similarity 100.0%; Pred. No. 6.7e-13;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 65
|||||
DB 49 CAGACTAGTATCCGCCGTCATCAATGACATCATCAGAGTACTGAGAGATGAATAGCT 106

RESULT 9

US-09-963-803-21

Sequence 21, Application US/09963803

Publication No. US20030028922A1

GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics

TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow

TITLE OF INVENTION: virus and cassava vein mosaic virus

FILE REFERENCE: 184332042

CURRENT APPLICATION NUMBER: US/09/963,803

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29

PRIOR APPLICATION NUMBER: PCT IB00/00370

PRIOR FILING DATE: 2000-10-05

NUMBER OF SEQ ID NOS: 39

SOFTWARE: PatentIn version 3.1

SEQ ID NO 21

LENGTH: 392

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: promoter MP1164
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(392)
OTHER INFORMATION:
US-09-963-803-21

Query Match 84.6%; Score 55; DB 9; Length 392;
Best Local Similarity 100.0%; Pred. No. 8,7e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 ACTAGTATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 20 ACTAGTATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74

RESULT 10
US-09-963-803-19

Sequence 19, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 19
LENGTH: 393
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1162
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(393)
OTHER INFORMATION:
US-09-963-803-19

Query Match 84.6%; Score 55; DB 9; Length 393;
Best Local Similarity 100.0%; Pred. No. 8,7e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 ACTAGTATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 20 ACTAGTATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74

RESULT 11
US-09-963-803-20

Sequence 20, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20

LENGTH: 462
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1163
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(462)
OTHER INFORMATION:
US-09-963-803-20

Query Match 84.6%; Score 55; DB 9; Length 462;
Best Local Similarity 100.0%; Pred. No. 9,1e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 ACTAGTATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 20 ACTAGTATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74

RESULT 12
US-09-963-803-22

Sequence 22, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 22
LENGTH: 600
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1165
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(600)
OTHER INFORMATION:
US-09-963-803-22

Query Match 84.6%; Score 55; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 9,9e-12;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 ACTAGTATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 65
Db 20 ACTAGTATCCCGCGTCATCATGACATCATCAGTACTGAGAGATGAATAGCT 74

RESULT 13
US-09-963-803-1

Sequence 1, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:

APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05

Qy 1 CATGCTGAGACTAGTATTCGCCCGTCATCATGATCATTCACAGTACTGAGGATGAA 60
Db 676 CTTTCTTACAGATCCTCTCCATCATCATCATCATCATCATCATTTATTTGACCAAGGCA 617
Qy 61 T 61

Search completed: May 11, 2003, 06:11:38
Job time : 67.3484 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 120.581 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-10

Perfect score: 62
Sequence: 1 gaagataagtcgtgatcgtg.....ctgaagtcggaatcgaag 62

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	62	21	AAA96844
2	62	100.0	305	19	AAV14022
3	62	100.0	317	21	AAA96837
4	62	100.0	371	21	AAA96839
5	62	100.0	392	19	AAV14019
6	62	100.0	392	21	AAA96855
7	62	100.0	393	21	AAA96853
8	62	100.0	411	19	AAV14021
9	62	100.0	420	19	AAV14026

10	62	100.0	462	21	AAA96854	Nucleotide sequenc
11	62	100.0	476	19	AAV14053	CesVWV promoter. C
12	62	100.0	515	21	AAA96836	Promoter from inte
13	62	100.0	515	22	AAF55505	Nucleotide sequenc
14	62	100.0	524	19	AAV14020	CesVWV promoter CYP
15	62	100.0	526	19	AAV14018	CesVWV promoter PA.
16	62	100.0	532	22	AAV11575	Cassava Vein Mosaic
17	62	100.0	593	22	AAF55507	Nucleotide sequenc
18	62	100.0	600	21	AAA96856	Nucleotide sequenc
19	62	100.0	653	22	AAF55506	Nucleotide sequenc
20	62	100.0	857	22	AAF55508	Nucleotide sequenc
21	62	100.0	931	22	AAF55509	Nucleotide sequenc
22	62	100.0	931	22	AAF55510	Nucleotide sequenc
23	62	100.0	931	22	AAF55510	Nucleotide sequenc
24	62	100.0	1839	24	AB157988	4-Hydroxyphenylpyr
25	62	100.0	4677	24	AB157989	4-Hydroxyphenylpyr
26	62	100.0	8187	24	AB158082	4-Hydroxyphenylpyr
27	62	100.0	8340	24	AB158082	4-Hydroxyphenylpyr
28	62	100.0	8340	24	AB158082	4-Hydroxyphenylpyr
29	62	100.0	12241	24	AB158082	4-Hydroxyphenylpyr
30	62	100.0	12241	24	AB158082	4-Hydroxyphenylpyr
31	56.2	90.6	482	19	AAV14027	Binary vector DNA
32	55.8	90.0	261	19	AAV14023	CesVWV promoter pde
33	50	80.6	301	21	AAA96841	CesVWV promoter pde
34	50	80.6	348	21	AAA96838	Nucleotide sequenc
35	50	80.6	398	21	AAA96840	Nucleotide sequenc
36	50	80.6	472	21	AAA96859	Nucleotide sequenc
37	50	80.6	541	21	AAA96857	Nucleotide sequenc
38	50	80.6	604	21	AAA96857	Nucleotide sequenc
39	40.6	65.5	441	19	AAV14033	CesVWV promoter pde
40	39.2	63.2	491	19	AAV14030	CesVWV promoter pde
41	26.6	42.9	735	22	AAK91933	Human CDNA 5'-end
42	26.2	42.3	1668	22	AAK91933	Human LI factor P4
43	26	41.9	833	22	AAK93287	Human CDNA clone r
44	26	41.9	939	22	ABA06449	Human CDNA SEQ ID
45	26	41.9	939	22	ABA06449	CDNA encoding nove

ALIGNMENTS

RESULT 1	AAA96844	AAA96844 standard; DNA; 62 BP.
ID	AAA96844	AAA96844
AC	AAA96844	AAA96844
DT	19-FEB-2001	(first entry)
XX	XX	XX
DE	Directional desoxynucleotide building block S3.	
KW	Promoter; intergenic region; Commelina yellow mottle virus;	
KW	chimeric expression promoter; plant vascular expression promoter;	
KW	plant green tissue expression promoter; Cassava vein mosaic virus;	
KW	transgenic plant; ss.	
XX	XX	XX
OS	Synthetic.	
XX	XX	XX
PN	WO200058485-A1.	
XX	XX	XX
PD	05-OCT-2000.	
XX	XX	XX
PF	29-MAR-2000; 2000MO-IB00370.	
XX	XX	XX
PR	29-MAR-1999; 99FR-0003925.	
XX	XX	XX
PA	(MERI-) MERISTEM THERAPEUTICS.	
XX	XX	XX
PI	Rance I, Gruber V, Theisen M;	
XX	XX	XX
DR	WPI; 2000-647238/62.	
XX	XX	XX
PT	Chimeric expression promoter for transgenic plant production, comprises	

```

PT Sequence from promoter comprising vascular expression region replaced
Pr with sequence from promoter comprising green tissue expression region
XX -
XX Disclosure; Page 23; 91pp; English.
XX
XX The present sequence represents a directional deoxynucleotide building
CC block, which was used to construct chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comelia yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
SQ
SQ Sequence 62 BP; 25 A; 4 C; 21 G; 12 T; 0 other;
Query Match 100.0%; Score 62; DB 21; Length 62;
Best Local Similarity 100.0%; Pred. No. 5e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps
OY 1 GAAGTAACGCTCGCATTTGGTGAAGAAGACAATAGAGCACACTTAGGTGAAAAATGTA 60
Db 1 GAAGTAAGAGTCGTGATTGTGAAGAGACATVAGAGCACAATGTAAAGTGAAAAATGTA 60
OY 61 AG 62
||
Db 61 AG 62
RESULT 2
AAV14022
ID AAV14022 standard; DNA; 305 BP.
XX
XX AAV14022;
AC 18-JUN-1998 (first entry)
XX
XX CbVMV promoter pc.
DE
XX
XX Cassava vein mosaic virus; CbVMV; promoter; cultivated crop;
KM tissue-specific expression control; transgenic plant; ss.
XX
XX Cassava vein mosaic virus.
OS
XX WO9748819-A1.
PM
XX
XX 24-DEC-1997.
PD
XX
XX 20-JUN-1997; 97WO-US10376.
PF
XX 20-JUN-1996; 96US-0020129.
PR
XX (SCRI ) SCRIPPS RES INST.
PA
XX
XX Beachy RN, De Kochko A, Fauquet C, Verdaguer B;
PI
XX
XX WPI; 1998-063157/06.
DR
XX
XX Cassava vein mosaic virus promoter - used to express heterologous
PT DNA sequences for producing transgenic plants having altered
PT phenotype(s).
CC
CC Claim 2; Page 77-78; 115pp; English.
XX
XX This sequence represents a cassava vein mosaic virus promoter, and
XX is a nucleic acid molecule of the invention. The promoter is capable of
XX initiating transcription of an operably linked heterologous nucleic acid
XX sequence in a plant cell. The CbVMV promoters are active in both monocot

```

CC and dicotyledonous species, and therefore can be readily applied to a
CC variety of cultivated crops. Although generally constitutive, the
CC derivative promoters include promoters that can regulate expression in a
CC tissue-specific manner, and therefore are useful for controlling
CC expression of heterologous genes in a tissue-specific manner. The
CC promoters can be used for producing transgenic plants with an altered
CC phenotype.

SQ Sequence 305 BP; 105 A; 46 C; 66 G; 88 T; 0 other;

Query Match 100.0%; Score 62; DB 19; Length 305;
Best Local Similarity 100.0%; Pred. No. 7e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 1 GAAGATTAAGCTCGGTGATTGGAAAGAACAACATACTGAAGTGAATAATGTA 60
|||||
Db 43 GAAGATTAAGCTCGGTGATTGGAAAGAACAACATACTGAAGTGAATAATGTA 102

Oy 61 AG 62
||
Db 103 AG 104

RESULT 3
AAA96837
ID AAA96837 standard; DNA; 317 BP.
AC AAA96837;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1116.
XX
KW Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
PD
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WQ-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERT-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Thelsen M;
DR WPI; 2000-647236/62.
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT -
XX
PS Claim 5; Page 81; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX Sequence 317 BP; 107 A; 61 C; 74 G; 75 T; 0 other;
 SQ Query Match 100.0%; Score 62; DB 21; Length 317;
 Best Local Similarity 100.0%; Pred.No. 7e-12;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGAATGTA 60
 DB 133 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGAATGTA 192

QY 61 AG 62
 DB 193 AG 194

RESULT 4
 ID AAA96839 standard; DNA; 371 BP.
 AC AAA96839;
 XX 19-FEB-2001 (first entry)
 DT Nucleotide sequence of chimeric expression promoter MP1146.
 DE Promoter; intergenic region; Commelina yellow mottle virus;
 XX chimeric expression promoter; plant vascular expression promoter;
 KM plant green tissue expression promoter; Cassava vein mosaic virus;
 KM transgenic plant; chimera; ss.
 XX Chimeric - Commelina yellow mottle virus.
 OS Chimeric - Cassava vein mosaic virus.
 XX WO200058485-A1.
 PN 05-OCT-2000.
 PD 29-MAR-2000; 2000WO-IB00370.
 XX 29-MAR-1999; 99FR-0003925.
 PR (MERI-) MERISTEM THERAPEUTICS.
 XX Rance I, Gruber V, Theisen M;
 PI WPI; 2000-647238/62.
 DR Chimeric expression promoter for transgenic plant production, comprises
 PT sequence from promoter comprising vascular expression region replaced
 PT with sequence from promoter comprising green tissue expression region
 PT
 XX Claim 5; Page 81; 91pp; English.
 PS The present sequence represents a chimeric promoter of the invention.
 CC The specification describes chimeric expression promoters. These
 CC chimeric promoters comprise a nucleic acid sequence which is derived
 CC from a first plant promoter, in which a plant vascular expression
 CC promoter region is replaced with a nucleic acid sequence derived from
 CC a second plant promoter comprising a plant green tissue expression
 CC promoter region. Preferably, the first plant promoter originates from
 CC Commelina yellow mottle virus, and the second plant promoter originates
 CC from the Cassava vein mosaic virus. Especially, the promoters are
 CC derived from intergenic regions. The chimeric promoters are useful
 CC for producing transgenic plants.
 XX Sequence 371 BP; 122 A; 68 C; 89 G; 92 T; 0 other;
 SQ Query Match 100.0%; Score 62; DB 21; Length 371;
 Best Local Similarity 100.0%; Pred.No. 7.3e-12;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGAATGTA 60
 DB 187 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGAATGTA 246

QY 61 AG 62
 DB 247 AG 248

RESULT 5
 ID AAV14019 standard; DNA; 392 BP.
 AC AAV14019;
 XX 18-JUN-1998 (first entry)
 DT CasMV promoter CVPL.
 DE Cassava vein mosaic virus; CasMV; promoter; cultivated crop;
 KM tissue-specific expression control; transgenic plant; ss.
 KM Cassava vein mosaic virus.
 OS WO9748819-A1.
 PN 24-DEC-1997.
 PD 20-JUN-1997; 97WO-US10376.
 XX 20-JUN-1996; 96US-0020129.
 PR (SCRI) SCRIPPS RES INST.
 PA Beachy RN, De Kochko A, Fauquet C, Verdaguer B;
 XX WPI; 1998-063157/06.
 DR Cassava vein mosaic virus promoter - used to express heterologous
 PT DNA sequences for producing transgenic plants having altered
 PT phenotype(s)
 XX Claim 2; Page 74; 115pp; English.
 PS This sequence represents a cassava vein mosaic virus promoter, and
 CC is a nucleic acid molecule of the invention. The promoter is capable of
 CC initiating transcription of an operably linked heterologous nucleic acid
 CC sequence in a plant cell. The CasMV promoters are active in both monocot
 CC and dicot plant species, and therefore can be readily applied to a
 CC variety of cultivated crops. Although generally constitutive, the
 CC derivative promoters include promoters that can regulate expression in a
 CC tissue-specific manner, and therefore are useful for controlling
 CC expression of heterologous genes in a tissue-specific manner. The
 CC promoters can be used for producing transgenic plants with an altered
 CC phenotype.
 XX Sequence 392 BP; 154 A; 64 C; 83 G; 91 T; 0 other;
 SQ Query Match 100.0%; Score 62; DB 19; Length 392;
 Best Local Similarity 100.0%; Pred.No. 7.3e-12;
 Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGAATGTA 60
 DB 185 GAAGATTAAGTCGGTGTGTTGTAAGAGACATAGAGACATGTAAAGTGAATGTA 244

QY 61 AG 62
 DB 245 AG 246

RESULT 6
 ID AAA96855

ID AAA96855 standard; DNA; 392 BP.
XX
AC AAA96855;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1164.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Cassava vein mosaic virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
XX
PS Claim 5; Page 86; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 392 BP; 127 A; 80 C; 87 G; 98 T; 0 other;
XX
Query Match 100.0%; Score 62; DB 21; Length 392;
Best Local Similarity 100.0%; Pred. No. 7.3e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GAAGATAAGTCGCGTGAATTGTGAAGACATAGAGACACATGTAGTGAATGTA 60
DB 135 GAAGATAAGTCGCGTGAATTGTGAAGACATAGAGACACATGTAGTGAATGTA 194
XX
QY 61 AG 62
DB 195 AG 196
XX
RESULT 7
ID AAA96853 standard; DNA; 393 BP.
XX
AC AAA96853;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1162.
XX

XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Cassava vein mosaic virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
XX
PS Claim 5; Page 85; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 393 BP; 128 A; 75 C; 93 G; 97 T; 0 other;
XX
Query Match 100.0%; Score 62; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 7.3e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 GAAGATAAGTCGCGTGAATTGTGAAGACATAGAGACACATGTAGTGAATGTA 60
DB 135 GAAGATAAGTCGCGTGAATTGTGAAGACATAGAGACACATGTAGTGAATGTA 194
XX
QY 61 AG 62
DB 195 AG 196
XX
RESULT 8
ID AAV14021 standard; DNA; 411 BP.
XX
AC AAV14021;
XX
DT 18-JUN-1998 (first entry)
XX
DE CeMVV promoter pB.
XX
KM Cassava vein mosaic virus; CeMVV; promoter; cultivated crop;
KM tissue-specific expression control; transgenic plant; ss.
XX
OS Cassava vein mosaic virus.
XX
PN WO9748819-A1.

XX WP1; 1998-063157/06.

PT Cassava vein mosaic virus promoter - used to express heterologous

PT DNA sequences for producing transgenic plants having altered

PT phenotype(s)

PS Claim 2; Page 80; 115pp; English.

XX This sequence represents a cassava vein mosaic virus promoter, and

CC is a nucleic acid molecule of the invention. The promoter is capable of

CC initiating transcription of an operably linked heterologous nucleic acid

CC sequence in a plant cell. The CaMV promoters are active in both monocot

CC and dicot plant species, and therefore can be readily applied to a

CC variety of cultivated crops. Although generally constitutive, the

CC derivative promoters include promoters that can regulate expression in a

CC tissue-specific manner, and therefore are useful for controlling

CC expression of heterologous genes in a tissue-specific manner. The

CC promoters can be used for producing transgenic plants with an altered

CC phenotype.

SX Sequence 420 BP; 149 A; 65 C; 91 G; 115 T; 0 other;

QY Query Match 100.0%; Score 62; DB 19; Length 420;

Db Best Local Similarity 100.0%; Pred. No. 7.4e-12;

QY Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGAATGTA 60

158 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGAATGTA 217

QY 61 AG 62

Db 218 AG 219

RESULT 10

AAA96854

ID AAA96854 standard; DNA; 462 BP.

XX AAA96854;

AC

XX 19-FEB-2001 (first entry)

DT Nucleotide sequence of chimeric expression promoter MP1163.

DE Promoter; intergenic region; Commelina yellow mottle virus;

XX chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

KM transgenic plant; chimera; ss.

XX

OS Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

XX WO200058485-A1.

XX 05-OCT-2000.

XX 29-MAR-2000; 2000WO-IB00370.

XX 29-MAR-1999; 99FR-0003925.

XX (MERI-) MERISTEM THERAPEUTICS.

XX Rance I, Gruber V, Theisen M;

XX WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 86; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX
XX Sequence 462 BP; 148 A; 87 C; 111 G; 116 T; 0 other;

XX
XX Query Match 100.0%; Score 62; DB 21; Length 462;
XX Best Local Similarity 100.0%; Pred. No. 7.6e-12;
XX Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
XX Db 135 GAAGATTAAGTCGGTGAATTGTGAAAGACATAGACACATGTAAAGTGAATGTA 60
XX |||
XX 61 AG 62
XX ||
XX Db 195 AG 196

XX
XX RESULT 11
XX AAV14053
XX ID AAV14053 standard; DNA; 476 BP.

XX
XX AAV14053:
XX 18-JUN-1998 (first entry)

XX
XX DE CaMV promoter.

XX
XX KM Cassava vein mosaic virus; CaMV; promoter; cultivated crop;
XX tissue-specific expression control; transgenic plant; ss.

XX
XX OS Cassava vein mosaic virus.

XX
XX PN WO9748819-A1.

XX
XX PD 24-DEC-1997.

XX
XX PF 20-JUN-1997; 97WO-US10376.

XX
XX PR 20-JUN-1996; 96US-0020129.

XX
XX PA (Scri) SCRIPPS RES INST.

XX
XX PI Beachy RN, De Kochko A, Fauquet C, Verdaguer B;

XX
XX DR WPI; 1998-063157/06.

XX
XX PT Cassava vein mosaic virus promoter - used to express heterologous
XX DNA sequences for producing transgenic plants having altered
XX phenotype(s)

XX
XX PS Disclosure; Page 87; 115pp; English.

XX
XX CC This sequence represents a cassava vein mosaic virus promoter, and
XX is a nucleic acid molecule of the invention. The promoter is capable of
XX initiating transcription of an operably linked heterologous nucleic acid
XX sequence in a plant cell. The CaMV promoters are active in both monocot
XX and dicot plant species, and therefore can be readily applied to a
XX variety of cultivated crops. Although generally constitutive, the
XX derivative promoters include promoters that can regulate expression in a
XX tissue-specific manner, and therefore are useful for controlling
XX expression of heterologous genes in a tissue-specific manner. The
XX promoters can be used for producing transgenic plants with an altered

CC phenotype.

XX
XX Sequence 476 BP; 188 A; 66 C; 110 G; 112 T; 0 other;

XX
XX Query Match 100.0%; Score 62; DB 19; Length 476;
XX Best Local Similarity 100.0%; Pred. No. 7.6e-12;
XX Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
XX Db 1 GAAGATTAAGTCGGTGAATTGTGAAAGACATAGACACATGTAAAGTGAATGTA 60
XX |||
XX 61 AG 62
XX ||
XX Db 324 AG 325

XX
XX RESULT 12
XX AAA96836
XX ID AAA96836 standard; DNA; 515 BP.

XX
XX AC AAA96836;
XX 19-FEB-2001 (first entry)

XX
XX DE Promoter from intergenic region of Cassava vein mosaic virus.

XX
XX KM Promoter; intergenic region; Commelina yellow mottle virus;
XX chimeric expression promoter; plant vascular expression promoter;
XX plant green tissue expression promoter; Cassava vein mosaic virus;
XX transgenic plant; ss.

XX
XX OS Cassava vein mosaic virus.

XX
XX PN WO200058485-A1.

XX
XX PD 05-OCT-2000.

XX
XX PF 29-MAR-2000; 2000WO-IB00370.

XX
XX PR 29-MAR-1999; 99FR-0003925.

XX
XX PA (MERI-) MERISTEM THERAPEUTICS.

XX
XX PI Rance I, Gruber V, Theisen M;

XX
XX DR WPI; 2000-647238/62.

XX
XX PT Chimeric expression promoter for transgenic plant production, comprises
XX sequence from promoter comprising vascular expression region replaced
XX with sequence from promoter comprising green tissue expression region

XX
XX PS Claim 4; Page 80; 91pp; English.

XX
XX CC The present sequence represents a promoter fragment from the intergenic
XX region of Cassava vein mosaic virus. The promoter is used to construct
XX chimeric expression promoters. These chimeric promoters comprise a
XX nucleic acid sequence which is derived from a first plant promoter,
XX in which a plant vascular expression promoter region is replaced with
XX a nucleic acid sequence derived from a second plant promoter comprising
XX a plant green tissue expression promoter region. Preferably, the first
XX plant promoter originates from Commelina yellow mottle virus, and the
XX second plant promoter originates from the Cassava vein mosaic virus.
XX The chimeric promoters are useful for producing transgenic plants.

XX
XX Sequence 515 BP; 198 A; 79 C; 109 G; 129 T; 0 other;

XX
XX Query Match 100.0%; Score 62; DB 21; Length 515;
XX Best Local Similarity 100.0%; Pred. No. 7.8e-12;
XX Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX
XX Db 1 GAAGATTAAGTCGGTGAATTGTGAAAGACATAGACACATGTAAAGTGAATGTA 60

DB 258 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 317
QY 61 AG 62
DB 318 AG 319

RESULT 13
ID AAF55505
AAF55505 standard; DNA; 515 BP.

AC AAF55505;

DT 29-MAY-2001 (first entry)

XX Nucleotide sequence of a region 5' to the CeMV translation start.

XX CeMV; enhancer domain; enhancer cassette; herbicide resistance;

KM fungal resistance; bacterial disease resistance; insect resistance;

XX plant ripening; plant degradation; plant colour; sweetness; ss.

OS Cassava vein mosaic virus.

PN WO200114573-A1.

PF 18-AUG-2000; 2000MO-US22595.

PR 19-AUG-1999; 99US-0149763.

PA (PROF-) PROFIGEN INC.

XX Xu D, Nielsen MT;

DR WPI; 2001-211307/21.

XX Enhancer cassette which forms an expression construct together with a

PT promoter useful for manipulating gene expression in plants, comprises

XX duplicated enhancer derived from cassava vein mosaic virus

PS Claim 3; Fig 1; 42pp; English.

XX The present sequence represents a region which is immediately 5' to

CC the Cassava vein mosaic virus (CeMV) translational start site. The

CC sequence is used as an enhancer domain in cassettes of the invention.

CC The specification describes an enhancer cassette comprising a duplicated

CC enhancer derived from CeMV. It can be operably linked to a nucleic

CC acid, the expression of which confers herbicide resistance, fungal,

CC bacterial disease resistance or insect resistance and regulates plant

CC ripening, degradation, colour and sweetness.

XX Sequence 515 BP; 198 A; 78 C; 110 G; 129 T; 0 other;

QY 1 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60

DB 258 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 317

QY 61 AG 62

DB 318 AG 319

RESULT 14
ID AAV14020
AAV14020 standard; DNA; 524 BP.

XX 18-JUN-1998 (first entry)

XX CeMV promoter CVP2.

XX Cassava vein mosaic virus; CeMV; promoter; cultivated crop;

KM tissue-specific expression control; transgenic plant; ss.

XX Cassava vein mosaic virus.

PN WO9748819-A1.

PD 24-DEC-1997.

PF 20-JUN-1997; 97MO-US10376.

PR 20-JUN-1996; 96US-0020129.

PA (SCRI) SCRIPPS RES INST.

PI Beachy RN, De Kochko A, Fauquet C, Verdaguer B;

DR WPI; 1998-063157/06.

XX Cassava vein mosaic virus promoter - used to express heterologous

PT DNA sequences for producing transgenic plants having altered

PS phenotype(s)

XX Claim 2; Page 75; 115pp; English.

XX This sequence represents a cassava vein mosaic virus promoter, and

CC is a nucleic acid molecule of the invention. The promoter is capable of

CC initiating transcription of an operably linked heterologous nucleic acid

CC sequence in a plant cell. The CeMV promoters are active in both monocot

CC and dicot plant species, and therefore can be readily applied to a

CC variety of cultivated crops. Although generally constitutive, the

CC derivative promoters include promoters that can regulate expression in a

CC tissue-specific manner, and therefore are useful for controlling

CC expression of heterologous genes in a tissue-specific manner. The

CC promoters can be used for producing transgenic plants with an altered

XX phenotype.

XX Sequence 524 BP; 201 A; 79 C; 112 G; 132 T; 0 other;

QY 1 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60

DB 262 GAAGATTAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 321

QY 61 AG 62

DB 322 AG 323

RESULT 15
ID AAV14018
AAV14018 standard; DNA; 526 BP.

AC AAV14018;

DT 18-JUN-1998 (first entry)

XX CeMV promoter PA.

XX Cassava vein mosaic virus; CeMV; promoter; cultivated crop;

XX tissue-specific expression control; transgenic plant; ss.

OS Cassava vein mosaic virus.

PN WO9748819-A1.

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 656.672 Seconds
(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-10

Perfect score: 62
Sequence: 1 gaagataagtcggtgctg.....tgaagtggaatgttaag 62

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBank: 1: gb_ba: 2: gb_ba: 3: gb_ba: 4: gb_ba: 5: gb_ba: 6: gb_ba: 7: gb_ba: 8: gb_ba: 9: gb_ba: 10: gb_ba: 11: gb_ba: 12: gb_ba: 13: gb_ba: 14: gb_ba: 15: gb_ba: 16: gb_ba: 17: gb_ba: 18: gb_ba: 19: gb_ba: 20: gb_ba: 21: gb_ba: 22: gb_ba: 23: gb_ba: 24: gb_ba: 25: gb_ba: 26: gb_ba: 27: gb_ba: 28: gb_ba: 29: gb_ba: 30: gb_ba: 31: gb_ba: 32: gb_ba: 33: gb_ba: 34: gb_ba: 35: gb_ba: 36: gb_ba: 37: gb_ba: 38: gb_ba: 39: gb_ba: 40: gb_ba: 41: gb_ba:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	100.0	62	AX036744	AX036744 Sequence
2	62	100.0	317	AX036737	AX036737 Sequence
3	62	100.0	371	AX036739	AX036739 Sequence
4	62	100.0	392	AX036755	AX036755 Sequence
5	62	100.0	333	AX036753	AX036753 Sequence
6	62	100.0	462	AX036754	AX036754 Sequence
7	62	100.0	515	AX036736	AX036736 Sequence
8	62	100.0	515	AX088388	AX088388 Sequence
9	62	100.0	532	AX202413	AX202413 Sequence
10	62	100.0	593	AX088390	AX088390 Sequence
11	62	100.0	600	AX036756	AX036756 Sequence
12	62	100.0	838	AX014764	AX014764 Sequence
13	62	100.0	853	AX088389	AX088389 Sequence
14	62	100.0	857	AX088391	AX088391 Sequence
15	62	100.0	931	AX088392	AX088392 Sequence
16	62	100.0	931	AX088393	AX088393 Sequence
17	62	100.0	931	AX088393	AX088393 Sequence
18	62	100.0	1036	AX014765	AX014765 Sequence
19	62	100.0	8158	CVU20341	CVU20341 Sequence
20	62	100.0	8159	CVU59751	CVU59751 Sequence
21	62	100.0	8340	AX329231	AX329231 Sequence
22	62	100.0	8340	AX338536	AX338536 Sequence
23	62	100.0	9285	AX093047	AX093047 Sequence
24	62	100.0	12241	AX412168	AX412168 Sequence
25	62	100.0	15077	AX093052	AX093052 Sequence
26	50	80.6	301	AX036741	AX036741 Sequence
27	50	80.6	348	AX036738	AX036738 Sequence
28	50	80.6	398	AX036740	AX036740 Sequence
29	50	80.6	472	AX036759	AX036759 Sequence
30	50	80.6	541	AX036758	AX036758 Sequence
31	50	80.6	604	AX036757	AX036757 Sequence
32	30	48.4	192667	AL845501	AL845501 Mus muscu
33	29.2	47.1	192740	AL773536	AL773536 Mus muscu
34	29	46.8	14000	AY029613	AY029613 Mus muscu
35	29	46.8	153483	AC068899	AC068899 Mus muscu
36	29	46.8	155587	AC022053	AC022053 Homo sapi
37	29	46.8	159927	AL353590	AL353590 Human DNA
38	29	46.8	180611	AL356284	AL356284 Human DNA
39	29	46.8	186758	AC073678	AC073678 Mus muscu
40	29	46.8	229700	AC073777	AC073777 Mus muscu
41	28.8	46.5	112537	AC112251	AC112251 Homo sapi
42	28.8	46.5	159173	AC117336	AC117336 Rattus no
43	28.8	46.5	184916	AC068998	AC068998 Mus muscu
44	28.8	46.5	222771	AC084069	AC084069 Mus muscu
45	28	45.2	132164	AC115137	AC115137 Rattus no

ALIGNMENTS

RESULT 1
LOCUS AX036744 62 bp DNA
DEFINITION Sequence 10 from Patent WO0058485.
ACCESSION AX036744
VERSION AX036744.1 GI:11226253
KEYWORDS
ORGANISM
SOURCE
synthetic construct.
synthetic construct.
artificial sequences.
REFERENCE 1 (bases 1 to 62)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 10 05-OCT-2000;

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source location/Qualifiers
1..62
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional desoxynucleotide building block S3"
BASE COUNT 25 a 4 c 21 g 12 t
ORIGIN
Query Match 100.0%; Score 62; DB 6; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 60
DB 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 60
QY 61 AG 62
DB 61 AG 62
RESULT 2
AX036737 317 bp DNA linear PAT 16-NOV-2000
LOCUS AX036737
DEFINITION Sequence 3 from Patent WO0058485.
ACCESSION AX036737
VERSION AX036737.1 GI:11226246
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 317)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 3 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source location/Qualifiers
1..317
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr116"
BASE COUNT 107 a 61 c 74 g 75 t
ORIGIN
Query Match 100.0%; Score 62; DB 6; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 60
DB 133 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 192
QY 61 AG 62
DB 193 AG 194
RESULT 3
AX036739 371 bp DNA linear PAT 16-NOV-2000
LOCUS AX036739
DEFINITION Sequence 5 from Patent WO0058485.
ACCESSION AX036739
VERSION AX036739.1 GI:11226248
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 371)

AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 5 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source location/Qualifiers
1..371
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1146"
BASE COUNT 122 a 68 c 89 g 92 t
ORIGIN
Query Match 100.0%; Score 62; DB 6; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 60
DB 187 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 246
QY 61 AG 62
DB 247 AG 248
RESULT 4
AX036755 392 bp DNA linear PAT 16-NOV-2000
LOCUS AX036755
DEFINITION Sequence 21 from Patent WO0058485.
ACCESSION AX036755
VERSION AX036755.1 GI:11226264
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 392)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow.
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 21 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source location/Qualifiers
1..392
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPr1164"
BASE COUNT 127 a 80 c 87 g 98 t
ORIGIN
Query Match 100.0%; Score 62; DB 6; Length 392;
Best Local Similarity 100.0%; Pred. No. 1.6e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 60
DB 135 GAAGTAAAGTCGGTGAATTGTGAAAGACATAGAGACACATGTAGTGAAGTGA 194
QY 61 AG 62
DB 195 AG 196
RESULT 5
AX036753 393 bp DNA linear PAT 16-NOV-2000
LOCUS AX036753
DEFINITION Sequence 19 from Patent WO0058485.
ACCESSION AX036753
VERSION AX036753.1 GI:11226262

Db 318 AG 319

RESULT 9
LOCUS AX0202413 532 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 1 from Patent WO0152620.
ACCESSION AX0202413
VERSION AX0202413.1 GI:15392159
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 532)
REFERENCE
AUTHORS Barbas, C.F., Steege, J.T., Guan, X. and Dalmida, B.
TITLE Methods and compositions to modulate expression in plants
JOURNAL Patent: WO 0152620-A 1 26-JUL-2001;
The Scripps Research Institute (US) ; SYNGENTA AGRICULTURAL
DISCOVERY, INC. (CA)
FEATURES
source location/Qualifiers
1..532
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Promoter CgMV"

BASE COUNT 204 a 82 c 111 g 135 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGAATGGAAGACATAGAGACACATGTAGGTGGAATAATGTA 60
|||||
Db 273 GAAGTAAAGTCGGTGAATGGAAGACATAGAGACACATGTAGGTGGAATAATGTA 332
|||||

QY 61 AG 62
Db 333 AG 334

RESULT 10
LOCUS AX088390 593 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114573.
ACCESSION AX088390
VERSION AX088390.1 GI:13397258
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 593)
REFERENCE
AUTHORS Xu, D. and Nielsen, M.T.
TITLE Duplicated cassava vein mosaic virus enhancers and uses thereof
JOURNAL Patent: WO 0114573-A 3 01-MAR-2001;
Profligen Inc. (US)
FEATURES
source location/Qualifiers
1..593
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="based on cassava vein mosaic virus"

BASE COUNT 228 a 86 c 135 g 144 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 593;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGAATGGAAGACATAGAGACACATGTAGGTGGAATAATGTA 60
|||||
Db 258 GAAGTAAAGTCGGTGAATGGAAGACATAGAGACACATGTAGGTGGAATAATGTA 317
|||||

QY 61 AG 62
Db 333 AG 334

Db 318 AG 319

RESULT 11
LOCUS AX036756 600 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 22 from Patent WO0058485.
ACCESSION AX036756
VERSION AX036756.1 GI:11226265
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 600)
REFERENCE
AUTHORS Rance, I., Theissen, M. and Gruber, V.
TITLE Chimeric expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 22 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
FEATURES
source location/Qualifiers
1..600
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPx1165"

BASE COUNT 188 a 111 c 147 g 154 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGAATGGAAGACATAGAGACACATGTAGGTGGAATAATGTA 60
|||||
Db 135 GAAGTAAAGTCGGTGAATGGAAGACATAGAGACACATGTAGGTGGAATAATGTA 194
|||||

QY 61 AG 62
Db 195 AG 196

RESULT 12
LOCUS AX014764 838 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 19 from Patent WO9953053.
ACCESSION AX014764
VERSION AX014764.1 GI:10041035
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 838)
REFERENCE
AUTHORS Lambers, M., Hofmann, J., Bulet, P. and Brookhart, G.L.
TITLE Gene coding for helminthine and use thereof
JOURNAL Patent: WO 9953053-A 19 21-OCT-1999;
LAMBERTY MIREILLE (FR); HOFMANN JULES (FR); BULET PHILIPPE (FR);
RHONE POULENC AGROCHIMIE (FR); BROOKHART GARY LEE (US)
FEATURES
source location/Qualifiers
1..838
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 296 a 134 c 175 g 233 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 838;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTGAATGGAAGACATAGAGACACATGTAGGTGGAATAATGTA 60
|||||

Db 264 GAAGATTAAGTCGGTGAATTTGAAAGACATAGAGACACATGTAAGTGGAAAAATGTA 323
QY 61 AG 62
Db 324 AG 325

RESULT 13

AX088389
LOCUS AX088389 853 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 2 from Patent WO0114573.
ACCESSION AX088389
VERSION AX088389.1 GI:13397257
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 853)
AUTHORS Xu,D. and Nielsen,M.T.
TITLE Duplicated cassava vein mosaic virus enhancers and uses thereof
JOURNAL Patent: WO 0114573-A 2 01-MAR-2001;
Profigen Inc. (US)
FEATURES
Location/Qualifiers
source 1. .853
/organism="synthetic construct"
/db xref="taxon:32630"
/note="based on cassava vein mosaic virus"

BASE COUNT 344 a 126 c 191 g 192 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 853;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGATTAAGTCGGTGAATTTGAAAGACATAGAGACACATGTAAGTGGAAAAATGTA 60
Db 596 GAAGATTAAGTCGGTGAATTTGAAAGACATAGAGACACATGTAAGTGGAAAAATGTA 655

QY 61 AG 62
Db 656 AG 657

RESULT 14

AX088391
LOCUS AX088391 857 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 4 from Patent WO0114573.
ACCESSION AX088391
VERSION AX088391.1 GI:13397259
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 857)
AUTHORS Xu,D. and Nielsen,M.T.
TITLE Duplicated cassava vein mosaic virus enhancers and uses thereof
JOURNAL Patent: WO 0114573-A 4 01-MAR-2001;
Profigen Inc. (US)
FEATURES
Location/Qualifiers
source 1. .857
/organism="synthetic construct"
/db xref="taxon:32630"
/note="based on cassava vein mosaic virus"

BASE COUNT 344 a 128 c 191 g 194 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 857;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGGTGAATTTGAAAGACATAGAGACACATGTAAGTGGAAAAATGTA 60
|||||

Db 522 GAAGATTAAGTCGGTGAATTTGAAAGACATAGAGACACATGTAAGTGGAAAAATGTA 581
QY 61 AG 62
Db 582 AG 583

RESULT 15

AX088392
LOCUS AX088392 931 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 5 from Patent WO0114573.
ACCESSION AX088392
VERSION AX088392.1 GI:13397260
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 931)
AUTHORS Xu,D. and Nielsen,M.T.
TITLE Duplicated cassava vein mosaic virus enhancers and uses thereof
JOURNAL Patent: WO 0114573-A 5 01-MAR-2001;
Profigen Inc. (US)
FEATURES
Location/Qualifiers
source 1. .931
/organism="synthetic construct"
/db xref="taxon:32630"
/note="based on cassava vein mosaic virus"

BASE COUNT 374 a 134 c 216 g 207 t
ORIGIN

Query Match 100.0%; Score 62; DB 6; Length 931;
Best Local Similarity 100.0%; Pred. No. 1.5e-10;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAAGATTAAGTCGGTGAATTTGAAAGACATAGAGACACATGTAAGTGGAAAAATGTA 60
Db 596 GAAGATTAAGTCGGTGAATTTGAAAGACATAGAGACACATGTAAGTGGAAAAATGTA 655

QY 61 AG 62
Db 656 AG 657

Search completed: May 11, 2003, 06:04:53
Job time : 658.672 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 24.6622 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-10

Perfect score: 62

Sequence: 1 gaagaaagtcgtgatg.....tgaagtggaatgtaag 62

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 44362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PTCUS COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	40.3	629	2	US-08-809-297-44
2	24.8	40.0	162450	4	US-09-345-882-1
3	24.4	39.4	2369	4	US-08-910-925-2
4	24	38.7	3454	4	US-08-961-527-207
5	23.6	38.1	130	1	US-08-182-175A-100
6	23.6	38.1	130	5	PCT-US92-06412-100
7	23.6	38.1	340	1	US-08-162-175A-104
8	23.6	38.1	340	5	PCT-US92-06412-104
9	23.6	38.1	2240	4	US-09-221-017B-687
10	23.4	37.7	599	2	US-08-809-297-43
11	23.2	37.4	464	4	US-08-098-327E-33
12	23.2	37.4	464	4	US-08-462-625-33
13	23.2	37.4	1001	4	US-09-641-618-575
14	23.2	37.4	1891	4	US-08-973-462-3
15	23.2	37.4	5361	4	US-08-973-462-2
16	23.2	37.4	6152	4	US-08-973-462-1
17	23	37.1	2910	4	US-09-206-942-31
18	23	37.1	2928	4	US-09-206-942-29
19	22.8	36.8	158	2	US-08-890-980-39
20	22.8	36.8	158	3	US-08-890-980-39
21	22.8	36.8	158	3	US-09-032-894-39
22	22.8	36.8	158	4	US-09-031-626-39
23	22.8	36.8	450	2	US-08-890-980-16
24	22.8	36.8	450	3	US-08-890-979-16
25	22.8	36.8	450	3	US-09-032-894-16
26	22.8	36.8	450	4	US-09-031-626-16
27	22.4	36.1	10962	2	US-08-816-155B-6

C 28	22.4	36.1	10962	3	US-09-079-587-6	Sequence 6, Appli
C 29	22.4	36.1	19877	2	US-08-816-155B-8	Sequence 8, Appli
C 30	22.4	36.1	19877	3	US-09-079-587-8	Sequence 8, Appli
C 31	22.4	36.1	168575	4	US-09-426-290-1	Sequence 1, Appli
32	22.2	35.8	851	4	US-09-171-209-61	Sequence 61, Appli
33	22.2	35.8	36159	4	US-09-749-588-3	Sequence 3, Appli
34	22	35.5	244	4	US-09-345-882-19	Sequence 19, Appli
35	22	35.5	6002	4	US-09-345-882-4	Sequence 4, Appli
36	21.8	35.2	480	6	5208144-34	Sequence 3, Appli
37	21.8	35.2	989	4	US-08-963-901-3	Sequence 3, Appli
38	21.8	35.2	2022	4	US-09-134-001C-1643	Sequence 1643, Ap
39	21.8	35.2	2400	4	US-08-963-901-1	Sequence 1, Appli
40	21.8	35.2	2400	4	US-08-963-901-5	Sequence 5, Appli
41	21.8	35.2	2515	1	US-08-061-465-3	Sequence 1, Appli
42	21.8	35.2	3033	3	US-08-836-325-1	Sequence 3, Appli
43	21.8	35.2	3131	3	US-09-035-648-23	Sequence 23, Appli
44	21.8	35.2	3131	4	US-09-001-951-23	Sequence 23, Appli
45	21.8	35.2	3131	4	US-08-818-829-23	Sequence 23, Appli

ALIGNMENTS

RESULT 1

US-08-809-297-44

Sequence 44, Application US/08809297

Patent No. 5948650

GENERAL INFORMATION:

APPLICANT: ARAKI, SHIGEKI

APPLICANT: TSUCHIYA, YOHICHI

TITLE OF INVENTION: GENETIC VARIETY IDENTIFYING METHOD IN

TITLE OF INVENTION: HOPS

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/809,297

FILING DATE: 06-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/02121

FILING DATE: 26-JUL-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP HEI 7-211328

FILING DATE: 28-JUL-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP HEI 8-130586

FILING DATE: 30-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24618

REFERENCE/DOCKET NUMBER: 2589-057-0PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 44:

SEQUENCE CHARACTERISTICS:

LENGTH: 629 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-809-297-44

Query March 40.3%; Score 25; DB 2; Length 629;
Best Local Similarity 63.4%; Pred. No. 3.6;
Matches 34; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 13 GGTGATTCTGCAAGACATAGACACATGTAGTGAATATGTA 61
DB 1 GCGCATTTCTGCAAGACACACACGACACAAATGTGATTAACATA 49

RESULT 2

US-09-345-882-1

Sequence 1, Application US/09345882

Patent No. 6399373

GENERAL INFORMATION:

APPLICANT: Bouguetel, Lydie

TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)

FILE REFERENCE: GENSET.031A

CURRENT APPLICATION NUMBER: US/09/345.882

PRIOR FILING DATE: 1999-06-30

PRIOR APPLICATION NUMBER: US 60/091,315

PRIOR FILING DATE: 1998-06-30

PRIOR APPLICATION NUMBER: US 60/111,909

NUMBER OF SEQ ID NOS: 140

SOFTWARE: Patent.pm

SEQ ID NO 1

LENGTH: 162450

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: allele

LOCATION: 72794

OTHER INFORMATION: 5-124-273 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 88073

OTHER INFORMATION: 5-127-261 : polymorphic base A or C

FEATURE:

NAME/KEY: allele

LOCATION: 90842

OTHER INFORMATION: 99-1437-325 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 93714

OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT

FEATURE:

NAME/KEY: allele

LOCATION: 97132

OTHER INFORMATION: 99-1442-224 : polymorphic base G or T

FEATURE:

NAME/KEY: allele

LOCATION: 97152

OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T

FEATURE:

NAME/KEY: allele

LOCATION: 99098

OTHER INFORMATION: 5-130-257 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 99117

OTHER INFORMATION: 5-130-276 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 103806

OTHER INFORMATION: 5-131-395 : polymorphic base A or T

FEATURE:

NAME/KEY: allele

LOCATION: 106940

OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 106940

OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 106940

OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 106940

OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 106940

OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 106106

OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 106149

OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT

FEATURE:

NAME/KEY: allele

LOCATION: 106308

OTHER INFORMATION: 5-135-357 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 106471

OTHER INFORMATION: 5-136-174 : polymorphic base C or T

FEATURE:

NAME/KEY: allele

LOCATION: 134134

OTHER INFORMATION: 5-140-120 : polymorphic base C or T

FEATURE:

NAME/KEY: allele

LOCATION: 134362

OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A

FEATURE:

NAME/KEY: allele

LOCATION: 134374

OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA

FEATURE:

NAME/KEY: allele

LOCATION: 146328

OTHER INFORMATION: 5-143-84 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 146345

OTHER INFORMATION: 5-143-101 : polymorphic base A or C

FEATURE:

NAME/KEY: allele

LOCATION: 150329

OTHER INFORMATION: 5-145-24 : polymorphic base A or G

FEATURE:

NAME/KEY: allele

LOCATION: 160031

OTHER INFORMATION: 5-148-352 : polymorphic base G or T

FEATURE:

NAME/KEY: allele

LOCATION: 72771..72817

OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30

FEATURE:

NAME/KEY: allele

LOCATION: 72771..72817

OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51

FEATURE:

NAME/KEY: allele

LOCATION: 88050..88096

OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31

FEATURE:

NAME/KEY: allele

LOCATION: 90819..90865

OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49

FEATURE:

NAME/KEY: allele

LOCATION: 90819..90865

OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

NAME/KEY: allele

LOCATION: 93690..93736

OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32

FEATURE:

LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:

Query Match 40.0%; Score 24.8; DB 4; Length 162450;
Best Local Similarity 61.3%; Pred. No. 16;
Matches 38; Conservative 1; Mismatches 23; Indels 0; Gaps 0;
1 GAAGATAGGTCGTGATTTGAAAGACATAGAGACACATGTAAAGTGAATGTA 60

|||||
Db 99061 GAAGAGAGAGAGAAATGATGAAAGAAAGAGAGGRTAATGCACTGAAGAGTGTA 99120
61 AG 62
99121 AG 99122

RESULT 3
US-08-910-925-2
Sequence 2, Application US/08910925
Patent No. 6162601
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN PININ SPLICER VARIANT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,925
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0365 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2369 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBROTO1
CLONE: 53219
US-08-910-925-2

Query Match 39.4%; Score 24.4; DB 4; Length 2369;
Best Local Similarity 63.8%; Pred. No. 8;
Matches 37; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

1 GAAGATAGGTCGTGATTTGAAAGACATAGAGACACATGTAAAGTGAATG 58
949 GAGGTAAAGTGTCTCAGCAGAGAGAAAGTGTGAGAGAGAGTAAATCAGACAAATG 1006
RESULT 4
US-08-961-527-207
Sequence 207, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: * Brooke, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 207:
SEQUENCE CHARACTERISTICS:
LENGTH: 3454 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-207

Query Match 38.7%; Score 24; DB 4; Length 3454;
Best Local Similarity 64.3%; Pred. No. 12;
Matches 36; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATGTAAGACATAGACACATGTAAGTGAAGA 56
DB 2616 GAAGTCAAGTGAAGGATTAATGCAAGAGAAATATGCTTAATGAAATGTAAGA 2671

RESULT 5
US-08-182-175A-100
Sequence 100, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:

NAME: Linda Axemethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 130 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DH5 alpha
FEATURE:
NAME/KEY: CDS
LOCATION: 3..116
OTHER INFORMATION: /function= "synthetic storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "sep"
OTHER INFORMATION: /standard_name= "SSP-seg5"
US-08-182-175A-100

Query Match 38.1%; Score 23.6; DB 1; Length 130;
Best Local Similarity 61.3%; Pred. No. 7;
Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 1 GAAGTAAAGTCGGTGAATGTAAGACATAGACACATGTAAGTGAAGA 60
DB 11 GAAGTGAAGAAAGCTCAAGAGGAATGCTAAGATGAAGCAAGAAATGGAAGAACTGA 70
QY 61 AG 62
DB 71 AG 72

RESULT 6
PCT-US92-06412-100
Sequence 100, Application PC/TUS9206412
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06412
FILING DATE: 19920807
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axemethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929

TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 100:
SEQUENCE CHARACTERISTICS:
LENGTH: 130 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DH5 alpha
FEATURE:
NAME/KEY: CDS
LOCATION: 3..116
OTHER INFORMATION: /function= "synthetic storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "asp"
OTHER INFORMATION: /standard_name= "SSP-seg5"
PCT-US92-06412-100

Query Match 38.1%; Score 23.6; DB 5; Length 130;
Best Local Similarity 61.3%; Pred. No. 7.7;
Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGCGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGAATGTA 60
DB 11 GAAGATTAAGTCAAGCTCAAGAGGAAATGCTTAAGATGAAGACGAAATGTGAAACTGAA 70

QY 61 AG 62
DB 71 AG 72

RESULT 7
US-08-182-175A-104
Sequence 104, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182.175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:

LENGTH: 340 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DH5 alpha
IMMEDIATE SOURCE:
CLONE: segment 534 [seg 534]
FEATURE:
NAME/KEY: CDS
LOCATION: 3..326
OTHER INFORMATION: /function= "synthetic seed storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "asp"
OTHER INFORMATION: /standard_name= "SSP-534"
US-08-182-175A-104

Query Match 38.1%; Score 23.6; DB 1; Length 340;
Best Local Similarity 61.3%; Pred. No. 9.7;
Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 GAAGATTAAGTCGCGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGAATGTA 60
DB 11 GAAGATTAAGTCAAGCTCAAGAGGAAATGCTTAAGATGAAGACGAAATGTGAAACTGAA 70

QY 61 AG 62
DB 71 AG 72

RESULT 8
PCT-US92-06412-104
Sequence 104, Application PC/TUS9206412
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06412
FILING DATE: 19920807
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 104:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: segment 534 (seg 534)
FEATURE:
NAME/KEY: CDS
LOCATION: 3..326
OTHER INFORMATION: /function= "synthetic seed storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "esp"
OTHER INFORMATION: /standard_name= "SSP-534"
PCT-US92-06412-104

Query Match 38.1%; Score 23.6; DB 5; Length 340;
Best Local Similarity 61.3%; Pred. No. 9.7;
Matches 38; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 GAAGTAAAGTCGGTATTGTAAGAGACATAGAGACACATGTAGTGAAGAAATGTA 60
DB 11 GAAGTGAAGAAAGCTCAAGAGAGAAATGCTTAGTGAAGACGAATGTGAAACTGAA 70

QY 61 AG 62
DB 71 AG 72

RESULT 9
US-09-221-017B-687/C
Sequence 687, Application US/09221017B
Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: ROSE, BRUCE C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD

CITY: PALO ALTO

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: IBM Compatible

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: P1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: P1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: P2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: MONROY, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-613-5600

TELEFAX: 706141

INFORMATION FOR SEQ ID NO: 687:
SEQUENCE CHARACTERISTICS:
LENGTH: 2240 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..2240
US-09-221-017B-687

Query Match 38.1%; Score 23.6; DB 4; Length 2240;
Best Local Similarity 64.8%; Pred. No. 15;
Matches 35; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2 AAGTAAAGTCGGTATTGTAAGAGACATAGAGACACATGTAAAGTGAAGA 55
DB 1405 AAAAATGTTGTTATTGCGACAGAGCCGAGACGCTACACGTGTCGGA 1352

RESULT 10
US-08-809-297-43

Sequence 43, Application US/08809297

Patent No. 594850

GENERAL INFORMATION:

APPLICANT: ARAKI, SHIGEKI

TITLE OF INVENTION: GENETIC VARIETY IDENTIFYING METHOD IN

NUMBER OF SEQUENCES: 48

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 SOUTH JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/809,297

FILING DATE: 06-MAY-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/02121

FILING DATE: 26-JUL-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP HEI 7-211328

FILING DATE: 28-JUL-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP HEI 8-130586

FILING DATE: 30-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24618

REFERENCE/DOCKET NUMBER: 2589-057-0PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 599 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-809-297-43

Query Match 37.7%; Score 23.4; DB 2; Length 599;
Best Local Similarity 67.3%; Pred. No. 13;
Matches 33; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 13 GGTATTGTGAAAGACATAGAGACATGTAAAGTGA 61
Db 1 GGCATTCTGCAAGACACACACGACAGAAATTGATAACATTA 49

RESULT 11

US-08-098-327E-33
Sequence 33, Application US/08098327E
Patent No. 6270771
GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUIHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 464 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-33

Query Match 37.4%; Score 23.2; DB 4; Length 464;
Best Local Similarity 61.7%; Pred. No. 14;
Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 1 GAAGATAGGTGCGTATTGTGAAGACATAGAGACACATGTAAAGTGA 60
Db 151 GAAGAAAAAGTTGAAGAAAGTGTGAAGAAATGCGAAGAAAGTGTGAAGAAATGTA 210

RESULT 12
US-08-462-625-33

Sequence 33, Application US/08462625
Patent No. 6319502
GENERAL INFORMATION:

APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUIHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/462,625
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/098,327

FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991

ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-078

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:
LENGTH: 464 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884

PUBLICATION DATE: 20-AUG-1992
US-08-462-625-33

Query Match 37.4%; Score 23.2; DB 4; Length 464;
Best Local Similarity 61.7%; Pred. No. 14;
Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 1 GAAGATAGGTGCGTATTGTGAAGACATAGAGACACATGTAAAGTGA 60
Db 151 GAAGAAAAAGTTGAAGAAAGTGTGAAGAAATGCGAAGAAAGTGTGAAGAAATGTA 210

RESULT 13

US-09-641-638-575
Sequence 575, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
APPLICANT: Cohen, Annick
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET 05ICPI
CURRENT APPLICATION NUMBER: US/09/641,638
CURRENT FILING DATE: 2000-08-16

PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 575
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 214
OTHER INFORMATION: 10-234-179 : deletion AA
NAME/KEY: misc_binding
LOCATION: 195..213
OTHER INFORMATION: 10-234-179.mis1
NAME/KEY: primer_bind
LOCATION: 36..56
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer_bind
LOCATION: 346..366
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_feature
LOCATION: 888
OTHER INFORMATION: n=a, g, c o r t
US-09-641-638-575

Query Match 37.4%; Score 23.2; DB 4; Length 1001;
Best Local Similarity 61.7%; Pred. No. 17;
Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
Qy 3 AGATTAAGTCGCGTATTGTGAAGACATAGAGACACATGTAAGTGAAGTGAAG 62
Db 511 AGCTGGGGCGAGAGAGAGAGAGAGACAGTAACTGAGTGAAGTGAAGTGAAG 570

RESULT 14
US-08-973-462-3
Sequence 3, Application US/08973462B
Patent No. 6191270
GENERAL INFORMATION:
APPLICANT: DRUIHE, PIERRE
APPLICANT: DAUBERSIES, PIERRE
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,462B
CURRENT FILING DATE: 1998-02-06
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007
EARLIER FILING DATE: 1995-06-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 1691
TYPE: DNA
ORGANISM: P. falciparum
FEATURE:
NAME/KEY: CDS
LOCATION: (2)..(1891)
US-08-973-462-3

Query Match 37.4%; Score 23.2; DB 4; Length 1891;
Best Local Similarity 61.7%; Pred. No. 20;
Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
Qy 1 GAAGTAAGTCGCGTATTGTGAAGACATAGAGACACATGTAAGTGAAGTGAAGTGA 60
Db 670 GAAGTAAGTCGCGTATTGTGAAGACATAGAGACACATGTAAGTGAAGTGAAGTGA 729

Db 320 GAAGTAAGTCGCGTATTGTGAAGACATAGAGACACATGTAAGTGAAGTGAAGTGA 379
RESULT 15
US-08-973-462-2
Sequence 2, Application US/08973462B
Patent No. 6191270
GENERAL INFORMATION:
APPLICANT: DRUIHE, PIERRE
APPLICANT: DAUBERSIES, PIERRE
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,462B
CURRENT FILING DATE: 1998-02-06
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007
EARLIER FILING DATE: 1995-06-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 5361
TYPE: DNA
ORGANISM: P. falciparum
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(5361)
US-08-973-462-2

Query Match 37.4%; Score 23.2; DB 4; Length 5361;
Best Local Similarity 61.7%; Pred. No. 25;
Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
Qy 1 GAAGTAAGTCGCGTATTGTGAAGACATAGAGACACATGTAAGTGAAGTGAAGTGA 60
Db 670 GAAGTAAGTCGCGTATTGTGAAGACATAGAGACACATGTAAGTGAAGTGAAGTGA 729

Search completed: May 11, 2003, 03:07:35
Job time : 53.8622 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 03:08:16 ; Search time 61.3784 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-10

Perfect score: 62

Sequence: 1 gaagataagtcggtgatcg.....tgaagtggaatgtaag 62

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	62	100.0	62	9	US-09-963-803-10	Sequence 10, Appl
2	62	100.0	317	9	US-09-963-803-3	Sequence 3, Appl1
3	62	100.0	371	9	US-09-963-803-5	Sequence 5, Appl1
4	62	100.0	332	9	US-09-963-803-21	Sequence 21, Appl
5	62	100.0	393	9	US-09-963-803-19	Sequence 19, Appl
6	62	100.0	462	9	US-09-963-803-20	Sequence 20, Appl
7	62	100.0	515	9	US-09-963-803-2	Sequence 2, Appl1
8	62	100.0	532	9	US-09-765-555-1	Sequence 1, Appl1
9	62	100.0	600	9	US-09-963-803-22	Sequence 22, Appl1
10	62	100.0	8340	10	US-09-847-057-4	Sequence 4, Appl1
11	62	100.0	8340	10	US-09-874-926-4	Sequence 4, Appl1
12	62	100.0	12241	12	US-10-033-190-5	Sequence 5, Appl1
13	50	80.6	301	9	US-09-963-803-7	Sequence 7, Appl1
14	50	80.6	348	9	US-09-963-803-4	Sequence 4, Appl1
15	50	80.6	398	9	US-09-963-803-6	Sequence 6, Appl1
16	50	80.6	472	9	US-09-963-803-25	Sequence 25, Appl
17	50	80.6	541	9	US-09-963-803-24	Sequence 24, Appl
18	50	80.6	604	9	US-09-963-803-23	Sequence 23, Appl
19	26	41.9	873	10	US-09-867-550-1223	Sequence 1223, Ap

20	26	41.9	939	9	US-09-860-670-34	Sequence 34, Appl
21	26	41.9	939	10	US-09-764-853-115	Sequence 115, Appl
22	26	41.9	990	9	US-10-160-501-12	Sequence 12, Appl
23	26	41.9	990	10	US-09-870-110-3	Sequence 3, Appl
24	26	41.9	1498	9	US-10-160-501-10	Sequence 10, Appl
25	26	41.9	1498	10	US-09-870-110-1	Sequence 1, Appl
26	26	41.9	1517	10	US-09-834-975-999	Sequence 999, Appl
27	25.4	41.0	672	10	US-09-770-149-403	Sequence 403, Appl
28	25.4	41.0	952	9	US-10-278-173-119	Sequence 119, Appl
29	25.2	40.6	3593	9	US-10-091-504-1634	Sequence 1634, Appl
30	25.2	40.6	3593	10	US-09-764-865-1634	Sequence 1634, Appl
31	24.6	39.7	2444	10	US-09-880-107-3030	Sequence 3030, Appl
32	24.4	39.4	1113	10	US-09-925-300-416	Sequence 416, Appl
33	24.4	39.4	1800	10	US-09-822-849A-1	Sequence 1, Appl
34	24	38.7	241	10	US-09-878-574-11995	Sequence 11995, A
35	24	38.7	422	9	US-09-818-995-17393	Sequence 17393, A
36	23.8	38.4	532	9	US-09-818-995-28694	Sequence 28694, A
37	23.6	38.1	530	10	US-09-764-898-59	Sequence 59, Appl
38	23.6	38.1	728	10	US-09-764-898-109	Sequence 109, Appl
39	23.6	38.1	1201	10	US-09-764-898-131	Sequence 131, Appl
40	23.6	38.1	640681	10	US-09-790-988-1	Sequence 1, Appl
41	23.4	37.7	556	9	US-09-764-891-1302	Sequence 1302, Appl
42	23.4	37.7	850	9	US-09-938-842A-3993	Sequence 3993, Appl
43	23.2	37.4	464	10	US-09-837-344-33	Sequence 33, Appl
44	23.2	37.4	1059	9	US-09-938-842A-2126	Sequence 2126, Appl
45	23.2	37.4	1569	10	US-09-974-300-2651	Sequence 2651, Appl

ALIGNMENTS

RESULT 1
US-09-963-803-10
Sequence 10, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
TITLE OF INVENTION: Virus and cassava vein mosaic virus
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963, 803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 62
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Directional desoxynucleotide building block
US-09-963-803-10

Query Match 100.0%; Score 62; DB 9; Length 62;
Best Local Similarity 100.0%; Pred. No. 2.6e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAGGTCGGTGTGTTGGAAGACATGAGACATGTAAGTGGGAAAGTGA 60
Db 1 GAAGATAGGTCGGTGTGTTGGAAGACATGAGACATGTAAGTGGGAAAGTGA 60

QY 61 AG 62
Db 61 AG 62

RESULT 2
US-09-963-803-3
Sequence 3, Application US/09963803.

Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 317
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1116
NAME/KEY: Promoter
LOCATION: (1)..(317)
OTHER INFORMATION:
US-09-963-803-3

Query Match 100.0%; Score 62; DB 9; Length 317;
Best Local Similarity 100.0%; Pred. No. 3.8e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 60
DB 133 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 192

QY 61 AG 62
DB 193 AG 194

RESULT 3
US-09-963-803-5
Sequence 5, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5
LENGTH: 371
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP1116
NAME/KEY: Promoter
LOCATION: (1)..(371)
OTHER INFORMATION:
US-09-963-803-5

Query Match 100.0%; Score 62; DB 9; Length 371;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 60

DB 187 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 246
QY 61 AG 62
DB 247 AG 248

RESULT 4
US-09-963-803-21
Sequence 21, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 21
LENGTH: 392
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP11164
NAME/KEY: Promoter
LOCATION: (1)..(392)
OTHER INFORMATION:
US-09-963-803-21

Query Match 100.0%; Score 62; DB 9; Length 392;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 60
DB 135 GAAGATAAGTCGGTGTGTAAGAGACATAGAGACACATGTAAAGTGAATGTA 194

QY 61 AG 62
DB 195 AG 196

RESULT 5
US-09-963-803-19
Sequence 19, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT FILING DATE: 2001-09-26
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 19
LENGTH: 393
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP11162

Query Match 100.0%; Score 62; DB 9; Length 393;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURE:
NAME/KEY: promoter
LOCATION: (1) - (393)
OTHER INFORMATION:
US-09-963-803-19

Query Match 100.0%; Score 62; DB 9; Length 393;
Best Local Similarity 100.0%; Pred. No. 4e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
|||
DB 135 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGAGACACATGTAAAGTGGAAAAATGTA 194

QY 61 AG 62
||
DB 195 AG 196

RESULT 6

US-09-963-803-20
Sequence 20, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
TITLE OF INVENTION: virus and cassava vein mosaic virus
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925;
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20
LENGTH: 462
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MPR1163
NAME/KEY: promoter
LOCATION: (1) - (462)
OTHER INFORMATION:
US-09-963-803-20

Query Match 100.0%; Score 62; DB 9; Length 462;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
|||
DB 135 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGAGACACATGTAAAGTGGAAAAATGTA 194

QY 61 AG 62
||
DB 195 AG 196

RESULT 7

US-09-963-803-2
Sequence 2, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
TITLE OF INVENTION: virus and cassava vein mosaic virus
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925

PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 515
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter from the intergenic region of Cassava Vein Mosaic virus
OTHER INFORMATION: of 515 bp in length EMBL
FEATURE:
NAME/KEY: promoter
LOCATION: (1) - (515)
OTHER INFORMATION:
US-09-963-803-2

Query Match 100.0%; Score 62; DB 9; Length 515;
Best Local Similarity 100.0%; Pred. No. 4.3e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
|||
DB 258 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 317

QY 61 AG 62
||
DB 318 AG 319

RESULT 8

US-09-765-555-1
Sequence 1, Application US/09765555
Publication No. US20030037355A1
GENERAL INFORMATION:
APPLICANT: The Scripps Research Institute
TITLE OF INVENTION: Methods and compositions to modulate
TITLE OF INVENTION: expression in plants
FILE REFERENCE: 27801-20014.40
CURRENT APPLICATION NUMBER: US/09/765,555
PRIOR FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 09/620,897
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 60/177,468
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 75
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 532
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter CAVV
US-09-765-555-1

Query Match 100.0%; Score 62; DB 9; Length 532;
Best Local Similarity 100.0%; Pred. No. 4.3e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 60
|||
DB 273 GAAGATAAGTCGGTGAATTGTGAAGAAGACATAGAGACACATGTAAAGTGGAAAAATGTA 332

QY 61 AG 62
||
DB 333 AG 334

RESULT 9

US-09-963-803-22
Sequence 22, Application US/09963803
Publication No. US20030028922A1

```

; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963, 803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP-1165
; NAME/KEY: promoter
; LOCATION: (1) (600)
; OTHER INFORMATION:
US-09-963-803-22

```

```

Query Match          100.0%; Score 62; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 4, 5e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY 1 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAGGTGGAATGTA 60
    |||
DB 135 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAGGTGGAATGTA 194
CY 61 AG 62
DB 195 AG 196

```

Handwritten: ECT 100

```

RESULT 10
US-09-847-057-4/C
; Sequence 4, Application US/09847057
; Patent No. US20020004943A1
; GENERAL INFORMATION:
; APPLICANT: AGRINOMICS, LLC.
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF A PAGODA PHENOTYPE IN PLANT
; FILE REFERENCE: PAGODA
; CURRENT APPLICATION NUMBER: US/09/847, 057
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 8340
; TYPE: DNA
; ORGANISM: Binary vector PAGI4002
US-09-847-057-4

```

```

Query Match          100.0%; Score 62; DB 10; Length 8340;
Best Local Similarity 100.0%; Pred. No. 8, 2e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY 1 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAGGTGGAATGTA 60
    |||
DB 7811 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAGGTGGAATGTA 7752

```

```

CY 61 AG 62
DB 7751 AG 7750

```

```

RESULT 11
US-09-874-926-4/C
; Sequence 4, Application US/09874926
; Patent No. US20020010950A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Agrinomics, LLC.
; TITLE OF INVENTION: Identification and characterization of a curly phenotype (cur) i.
; FILE REFERENCE: curly
; CURRENT APPLICATION NUMBER: US/09/874, 926
; CURRENT FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 8340
; TYPE: DNA
; ORGANISM: PAGI4002
US-09-874-926-4

```

```

Query Match          100.0%; Score 62; DB 10; Length 8340;
Best Local Similarity 100.0%; Pred. No. 8, 2e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY 1 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAGGTGGAATGTA 60
    |||
DB 7811 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAGGTGGAATGTA 7752
CY 61 AG 62
DB 7751 AG 7750

```

Handwritten: ECT 100

```

RESULT 12
US-10-033-190-5/C
; Sequence 5, Application US/10033190
; Patent No. US20020133848A1
; GENERAL INFORMATION:
; APPLICANT: Exelixis Plant Sciences, Inc.
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF AN ANTHOCYANIN MUTANT (AN
; FILE REFERENCE: EP01-002C
; CURRENT APPLICATION NUMBER: US/10/033, 190
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: US 60/244, 685
; PRIOR FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 12241
; TYPE: DNA
; ORGANISM: PAG2370
US-10-033-190-5

```

Handwritten: ECT 100

```

Query Match          100.0%; Score 62; DB 12; Length 12241;
Best Local Similarity 100.0%; Pred. No. 9e-12;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

CY 1 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAGGTGGAATGTA 60
    |||
DB 3217 GAAGATAAGGTCGGTGAATTGTGAAAGAGACATAGAGACACATGTAGGTGGAATGTA 3158

```

```

CY 61 AG 62
DB 3157 AG 3156

```

```

RESULT 13
US-09-963-803-7
; Sequence 7, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963, 803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925

```

; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Promoter MP1154
; NAME/KEY: promoter
; LOCATION: (1)..(301)
; OTHER INFORMATION:
US-09-963-803-7

Query Match 80.6%; Score 50; DB 9; Length 301;
Best Local Similarity 98.4%; Pred. No. 4.9e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACATGTAAAGTGAATAATGTA 60
DB 120 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACATGTAAAGTGAATAATGTA 178

QY 61 AG 62
DB 179 AG 180

RESULT 14

US-09-963-803-4
; Sequence 4; Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 348
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Promoter MP1117
; NAME/KEY: promoter
; LOCATION: (1)..(348)
; OTHER INFORMATION:
US-09-963-803-4

Query Match 80.6%; Score 50; DB 9; Length 348;
Best Local Similarity 98.4%; Pred. No. 5.1e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACATGTAAAGTGAATAATGTA 60
DB 167 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACATGTAAAGTGAATAATGTA 225

QY 61 AG 62
DB 226 AG 227

RESULT 15

US-09-963-803-6
; Sequence 6; Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Promoter MP1147
; NAME/KEY: promoter
; LOCATION: (1)..(398)
; OTHER INFORMATION:
US-09-963-803-6

Query Match 80.6%; Score 50; DB 9; Length 398;
Best Local Similarity 98.4%; Pred. No. 5.3e-08;
Matches 61; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACATGTAAAGTGAATAATGTA 60
DB 217 GAAGTAAAGTCGGTGTGTAAGAGACATAGAGACATGTAAAGTGAATAATGTA 275

QY 61 AG 62
DB 276 AG 277

Search completed: May 11, 2003, 06:11:41
Job time : 64.3784 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 996.662 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-10

Perfect score: 62
Sequence: 1 gaagataagctcgctgctg.....tgaagctggaataatgaag 62

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_estbma:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gsa:*
18: em_gsa_hum:*
19: em_gsa_inv:*
20: em_gsa_pln:*
21: em_gsa_vit:*
22: em_gsa_fun:*
23: em_gsa_mam:*
24: em_gsa_mus:*
25: em_gsa_other:*
26: em_gsa_pro:*
27: em_gsa_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	29	46.8	699	AQ156245
2	28.8	46.5	359	BG278522
3	28.8	46.5	548	BG278526
4	28.8	46.5	576	A1399593
5	28.8	46.5	585	BG280606
6	28	45.2	214	AZ484766

Result No.	Score	Query Match Length	ID	Description
7	28	45.2	383	AQ165808
8	27.2	43.5	408	AQ212331
9	27.2	43.5	417	BG279968
10	27.2	43.5	670	BG278034
11	26.8	43.2	656	AZ570132
12	26.8	42.9	397	AQ410127
13	26.6	42.9	453	AQ620909
14	26.6	42.9	602	FR0020027
15	26.6	42.9	649	CNS03MWM
16	26.4	42.6	476	BF622523
17	26.4	42.6	501	BM360131
18	26.4	42.6	550	AJ435435
19	26.4	42.6	564	AZ827513
20	26.4	42.6	598	BQ458384
21	26.4	42.6	737	BE602510
22	26.4	42.6	243	N70246
23	26.4	41.9	421	B14130
24	26.4	41.9	438	AQ350013
25	26.4	41.9	458	AA442502
26	26.4	41.9	500	AA436699
27	26.4	41.9	514	AQ438893
28	26.4	41.9	561	AW958726
29	26.4	41.9	571	AQ687904
30	26.4	41.9	574	BF590238
31	26.4	41.9	592	BF109143
32	26.4	41.9	598	FR0034799
33	26.4	41.9	615	BF055540
34	26.4	41.9	619	BH446500
35	26.4	41.9	625	CNS06009
36	26.4	41.9	631	AV697865
37	26.4	41.9	634	BG699782
38	26.4	41.9	678	BE269715
39	26.4	41.9	678	BU002041
40	26.4	41.9	683	BG281521
41	26.4	41.9	700	BG716815
42	26.4	41.9	705	BE269319
43	26.4	41.9	705	BE269605
44	26.4	41.9	715	BU008674
45	26.4	41.9	725	BQ114742

ALIGNMENTS

RESULT 1
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AQ156245
nxbdb005D02r
AQ156245
AQ156245.1
GI:3592027
699 bp
DNA
linear
GSS
12-SEP-1998
nxbdb005D02r
CGI Rice BAC Library
Oryza sativa genomic clone
Oryza sativa
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 699)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
On Sep 10, 1998 this sequence version replaced gi:3552334.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seg primer: GGAACGCTATGACCATG
Class: BAC ends
High quality sequence stop: 394.
Location/Qualifiers

```

SOURCE
1..699
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipondare"
/db xref="taxon:4530"
/clone="nbx0005D02r"
/clone_1lb="CUGI Rice BAC library"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/notes="Vector: pBeloBAC11, Site 1: HindIII, Site 2: HindIII, Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipondare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT      163 a      250 c      138 g      145 t      3 others
ORIGIN

Query Match      46.8%; Score 29; DB 17; Length 699;
Best Local Similarity 67.2%; Pval: 1.1e-02;
Matches 41; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Oy 1 GAAGTAAAGCTGGTGCATTTGTAAGAGACATAGAGACATGTAAAGTGAATAATGTA 60
    |||||
Db 612 GAAGAGAGATTGTTAAGGATTAAAGATGTTGTTAGACACCTGTGAAGGGGGAATATATC 671
Oy 61 A 61
    |
Db 672 A 672

RESULT 2
BG278522
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

FEATURES	SOURCE	Location/Qualifiers
1. 359		
/organism="Neurospora crassa"		
/strain="wild type"		
/db_xref="taxon:5141"		
/clone="asc01np"		
/clone_id="Neurospora crassa sexual cDNA library, Uni-zap vector system"		
/issue_type="perithecia (fruiting bodies)"		
/dev_stage="sexual"		
/lab_host="E. coli strain SOLR"		
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript ; 3' end of cDNA cloned into XhoI site of pBluescript"		
BASE COUNT	127 a 78 c 97 g 57 t	
ORIGIN		
Query Match	46.5%; Score 28.8; DB 12; Length 359;	
Best Local Similarity	75.0%; Pred. No. 1e+02;	
Matches	36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;	
5 ATAAAGTCGCTGATTGTGAAAGACATAGAGACACATGTAAGCTGG 52		
224 ACAAGCGAGTACACGAGAGAAAGAAATTAAGAGACAGTAAAGGGGG 271		
548 bp mRNA		
b3g07np.f1 Neurospora crassa sexual cDNA library, Uni-zap vector		
system Neurospora crassa cDNA clone b3g07np 3', mRNA sequence.		
Accession	BG279526	
Version	BG279526.1	GI:13076980
Keywords	EST.	
Source	Neurospora crassa	
Organism	Neurospora crassa	
Reference	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariales; Sordariaceae; Neurospora.	
Authors	1 (bases 1 to 548)	
Title	Kupfer, D., Lai, H., Nelson, M. and Roe, B.	
Journal	ESTs from a Neurospora crassa Sexual cDNA Library	
Comment	Unpublished (2001)	
Other ESTs	b3g07np.r1	
Contact	Bruce A. Roe, University of Oklahoma, broeou.edu	
Department of Chemistry and Biochemistry		
Advanced Center for Genome Technology, University of Oklahoma		
620 Parrington Oval, Norman, OK 73019, USA		
Tel: 405 325 4912		
Fax: 405 325 7762		
Email: broeou.edu		
Contact Dr. Mary Anne Nelson, Department of Biology, University of New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu) regarding clone availability		
Seq primer: M13 Universal Forward Primer		
High quality sequence stop: 442.		
Location/Qualifiers		
1. 548		
/organism="Neurospora crassa"		
/strain="wild type"		
/db_xref="taxon:5141"		
/clone="b3g07np"		
/clone_id="Neurospora crassa sexual cDNA library, Uni-zap vector system"		
/issue_type="perithecia (fruiting bodies)"		
/dev_stage="sexual"		
/lab_host="E. coli strain SOLR"		
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript ; 3' end of cDNA cloned into XhoI site of pBluescript"		
BASE COUNT	182 a 146 c 123 g 97 t	
ORIGIN		
Query Match	46.5%; Score 28.8; DB 12; Length 548;	

Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 ATAAAGTCGGATTGTGAAGAAGACATAGAGACACATGTAAGTGG 52
DB 234 ACACGGCGACGTGACAGAGAAAGAAATTAAGAGACAGTAAAGGGG 281

RESULT 4
AI399593

LOCUS AI399593 576 bp mRNA linear EST 08-FEB-1999
DEFINITION NCP6677 Subtracted Perithecial Neurospora crassa cDNA clone SP656

ACCESSION AI399593

VERSION AI399593.1 GI:4242680

KEYWORDS

SOURCE

ORGANISM

Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

REFERENCE 1 (bases 1 to 576)
Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,

Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueyes, E.,
Cushing, T., Bretz, A., Fleharty, M., Gorman, M., Judson, K., Miller, R.,
Ortega, J., Pavlova, I., Perez, J., Todisco, S., Trujillo, R.,
Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S. and Natvig,
D.O.

Expressed sequences from conidial, mycelial, and sexual stages of
Neurospora crassa

JOURNAL Fungal Genet. Biol. 21, 348-363 (1997)

COMMENT Contact: Natvig, D.O./Nelson, M.A.
Department of Biology
University of New Mexico
Casceter Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304
Email: ngpbiology.unm.edu

FEATURES
source location/Qualifiers

1..576

/organism="Neurospora crassa"

/strain="fl a (FGSC 4347)"

/db_xref="taxon:5141"

/clone="SP656"

/clone_lib="Subtracted Perithecial"

/sex="Mating type a (fluffy), fertilized"

/tissue_type="Perithecia"

/dev_stage="Fruiting Body"

/lab_host="E. coli"

/note="Vector: pBluescript SK (-); Site 1: EcoRI; Site 2:
XhoI; mRNA isolated from 5 day old perithecia (fruiting
bodies) of the fluffy strain fl a (Mating type a),
fertilized with conidia from 74-OR23-IV A (Mating type
A). cDNA directionally cloned into pBluescript SK (-)
using the Uni-ZAP XR vector system (Stratagene, La Jolla,
CA). Previously identified highly expressed clones were
subtracted from this library."

BASE COUNT 168 a 144 c 125 g 139 t
ORIGIN

Query Match 46.5%; Score 28.8; DB 9; Length 576;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 ATAAAGTCGGATTGTGAAGAAGACATAGAGACACATGTAAGTGG 52
DB 101 ACACGGCGACGTGACAGAGAAAGAAATTAAGAGACAGTAAAGGGG 148

RESULT 5
BG280606 585 bp mRNA linear EST 21-FEB-2001
LOCUS BG280606
DEFINITION c5b03mp.fl Neurospora crassa sexual cDNA library, uni-zap vector

ACCESSION BG280606
VERSION BG280606.1 GI:13079182
KEYWORDS
SOURCE
ORGANISM

Neurospora crassa.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.

REFERENCE 1 (bases 1 to 585)

Kupfer, D., Lai, H., Nelson, M. and Roe, B.

ESTs from a Neurospora crassa Sexual cDNA Library

Unpublished (2001)

Other ESTs: c5b03mp.r1

Contact: Bruce A. Roe, University of Oklahoma, broeou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Farrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broeou.edu

Contact Dr. Mary Anne Nelson, Department of Biology, University of

New Mexico, Albuquerque, NM 87131 (e-mail address manelsonunm.edu)

regarding clone availability

Seq primer: M13 Universal Forward Primer

High quality sequence stop: 474.

FEATURES
source location/Qualifiers

1..585

/organism="Neurospora crassa"

/strain="wild type"

/db_xref="taxon:5141"

/clone="c5b03mp"

/clone_lib="Neurospora crassa sexual cDNA library, Uni-zap
vector system"

/tissue_type="perithecia (fruiting bodies)"

/dev_stage="sexual"

/lab_host="E. coli strain SOLR"

/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 187 a 161 c 126 g 111 t
ORIGIN

Query Match 46.5%; Score 28.8; DB 12; Length 585;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 36; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 ATAAAGTCGGATTGTGAAGAAGACATAGAGACACATGTAAGTGG 52
DB 224 ACACGGCGACGTGACAGAGAAAGAAATTAAGAGACAGTAAAGGGG 271

RESULT 6
AZ484766/c 214 bp DNA linear GSS 05-OCT-2000
LOCUS AZ484766/c
DEFINITION IM0311P23F Mouse 10kb plasmid UUCGM library Mus musculus genomic
clone UUCGM0311P23 F, DNA sequence.

ACCESSION AZ484766

VERSION AZ484766.1 GI:10649926

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 214)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss
University of Utah Genome Center

University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 1000 Std Error: 0.00
Plate: 0311 row: P column: 23
Seq primer: CCTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 214.
Location/Qualifiers:
1. .214
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0311P23"
/clone_1lb="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g14732114[g14732114]_1) a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 52 a 51 c 39 g 72 t
ORIGIN

Query Match 45.2%; Score 28; DB 17; Length 214;
Best Local Similarity 71.2%; Pred. No. 1.5e+02;
Matches 37; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 2 AAGATAGCTGGTATGTGAAAGACATAGACACATGTAAGTGA 53
Db 103 AAGACAAAGTACTGCTGAGGAGAGACCTTAGAGACACATGATTTTGA 52

RESULT 7
LOCUS A0165808/c 383 bp DNA linear GSS 16-OCT-1998
DEFINITION HS_3071_A1_G12_MR_CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3071 Col=23 Row=M, DNA sequence.
ACCESSION A0165808
VERSION A0165808.1 GI:3564003
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 383)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL Mahairas GG, Wallace JC, Hood L
MEDLINE 99380589
COMMENT High Throughput Sequencing Center

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3071 row: M column: 23
Class: BAC ends
High quality sequence stop: 383.
Location/Qualifiers:
1. .383
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3071 Col=23 Row=M"
/clone_1lb="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 92 a 98 c 86 g 105 t 2 others
ORIGIN

Query Match 45.2%; Score 28; DB 17; Length 383;
Best Local Similarity 71.2%; Pred. No. 1.5e+02;
Matches 37; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 7 AAGTCCGTATGTGAAAGACATAGACACATGTAAGTGAATG 58
Db 169 ATGAAAGTACTGTGAAAGGAGAGACACACAGCTGTGAAATG 118

RESULT 8
LOCUS A0212331 418 bp DNA linear GSS 18-SEP-1998
DEFINITION HS_2240_A2_G10_MR_CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=2240 Col=20 Row=M, DNA sequence.
ACCESSION A0212331
VERSION A0212331.1 GI:3623532
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 418)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL Mahairas GG, Wallace JC, Hood L
MEDLINE 99380589
COMMENT High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 2240 row: M column: 20
Class: BAC ends
High quality sequence stop: 418.
Location/Qualifiers:
1. .418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=2240 Col=20 Row=M"
/clone_1lb="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in
E-Coli DH10B"

BASE COUNT 129 a 63 c 102 g 124 t
ORIGIN

Query Match 45.2%; Score 28; DB 17; Length 418;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 40; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 1 GAAGTAAGCGGTGATTTGTAAGAGACATAGAGACATGTGAAGTGAAGTA 60
 DB 176 GGAGATAAAATTTGGACATTTGAAATAGACAGTGGATGAGATGAGAGAGAGTA 235

RESULT 9
 BG279968 407 bp mRNA linear EST 21-FEB-2001
 LOCUS b7h08np.f1 Neurospora crassa sexual cDNA library, Uni-zap vector
 DEFINITION system Neurospora crassa cDNA clone b7h08np 3', mRNA sequence.
 ACCESSION BG279968
 VERSION BG279968.1 GI:13077915
 KEYWORDS EST.
 SOURCE Neurospora crassa.
 ORGANISM Neurospora crassa
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariales; Sordariaceae; Neurospora.

REFERENCE
 1 (bases 1 to 407)
 AUTHORS Kupfer, D., Lai, H., Nelson, M. and Roe, B.
 TITLES ESTs from a Neurospora crassa Sexual cDNA Library
 JOURNAL Unpublished (2001)
 COMMENT Other ESTs: b7h08np.r1
 Contact: Bruce A. Roe, University of Oklahoma, broeou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broeou.edu
 Contact: Dr. Mary Anne Nelson, Department of Biology, University of
 New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu)
 regarding clone availability
 Seq primer: M13 Universal Forward Primer
 High quality sequence stop: 402.

FEATURES
 source
 1..407
 /organism="Neurospora crassa"
 /strain="wild type"
 /db_xref="taxon:5141"
 /clone="b7h08np"
 /clone_lib="Neurospora crassa sexual cDNA library, Uni-zap
 vector system"
 /tissue_type="perithecia (fruiting bodies)"
 /dev_stage="sexual"
 /lab_host="E. coli strain SOLR"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
 XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
 ; 3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 136 a 110 c 85 g 76 t

ORIGIN

Query Match 43.9%; Score 27.2; DB 12; Length 407;
 Best Local Similarity 72.9%; Pred. No. 3.2e+02;
 Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 5 ATAAAGTCGATGATTTGTAAGAGACATAGAGACATGTGAAGTGG 52
 DB 73 ACACGGGAGTAGACAGAGAAAGAAATTAAGAGACAGTAGGGCGG 120

RESULT 10
 BE728034 670 bp mRNA linear EST 15-SEP-2000
 LOCUS 601561322F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:383092 5',
 DEFINITION mRNA sequence.
 ACCESSION BE728034
 VERSION BE728034.1 GI:10142026
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (bases 1 to 670)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC/DCPD/DP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
 Plate: LLCMS05 row: 1 column: 05
 High quality sequence stop: 670.

FEATURES
 source
 1..670
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3830932"
 /clone_lib="NIH_MGC_20"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACTGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 179 a 142 c 189 g 160 t

ORIGIN

Query Match 43.9%; Score 27.2; DB 12; Length 670;
 Best Local Similarity 67.9%; Pred. No. 3.9e+02;
 Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 3 AGATAAGTCGATGATTTGTAAGAGACATAGAGACATGTGAAGTGA 58
 DB 371 AGGAGAAGTTGACAGTTGTCTTAAGACATAGACACATCAAGTGAAGT 426

RESULT 11
 A2570132/c 656 bp DNA linear GSS 15-MAY-2001
 LOCUS 270PG10 PV MBN #30 Plasmodium vivax genomic 3', DNA sequence.
 DEFINITION A2570132
 ACCESSION A2570132
 VERSION A2570132.1 GI:13980913
 KEYWORDS GSS.
 SOURCE malaria parasite P. vivax.
 ORGANISM Plasmodium vivax
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE
 1 (bases 1 to 656)
 AUTHORS Carlson, J.M.-R. and Dame, J.B.
 TITLES The Plasmodium vivax and P. berghei gene sequence tag projects
 JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
 COMMENT Contact: Dame JB
 Dept. of Pathobiology, College of Veterinary Medicine
 University of Florida
 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
 Tel: 352 392 4700
 Fax: 352 392 9704
 Email: damej@mail.vetmed.ufl.edu
 Seq primer: M13(-20) forward
 Class: shotgun.

FEATURES
 source
 1..656
 /organism="Plasmodium vivax"
 /strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
 497-598)"

```

BASE COUNT      170 a      140 c      146 g      199 t      1 others
ORIGIN
/dB xref="taxon:5855"
/clone_id="Pv MBV #30"
/dev_stage="asexual blood forms"
/lab_host="Salmixi boliviensis"
/note="vector: pBluescript SK(+)"
/excised from lambda ZAP: site 1: Ecor V; site 2: Ecor V;
Host leukocytes were extracted from P. vivax infected
blood using the following methods: first, infected blood
was activated by the addition of 0.5 ml of ADP (40mg/ml)
per 10 ml blood. Then blood was passed over a column of
acid washed 0.1 mm glass beads, then through a Plasmidipure
filter, followed by passage through a column of pre-wet
Pharmacia Cpl powder (1:2 ratio volume of blood to Cpl),
and finally centrifuged through a 50% Percoll density
cushion. Purified DNA was digested with mung bean nucleases
in the presence of 44% formamide at 500C as described
(Venick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). Digested DNA was
blunt-ended using T4 DNA polymerase and size fractionated
over a Sepharose CL-2B column. Fractions in the size ranges
500bp-4kb were ligated into the Eco RV site of pBluescript
SK(+), and E. coli XL-10 Gold transformed with the
ligation mixture."

```

Query Match	43.2%	Score 26.8	DB 17	Length 656
Best Local Similarity	68.5%	Pred. No. 5	1e+02	
Matches 37	Conservative 0	Mismatches 17	Indels 0	Gaps 0

Oy	7	AAGTCGCGATGTTGAAAAGACATAGACACATGTAGCTGAAAATGTA	60
Db	155	AAGTACGCTTTTATAGAAAGATATAGGAAAACTGAAAAGGGAATTTA	102

RESULT 12	
AQ410127	
LOCUS	AQ410127
DEFINITION	HS_5083_BJ_A05 SP6E RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate=659 Col=9 Row=B, DNA sequence.

ORGANISM	REFERENCE
Homo sapiens	
Euryarchaeota; Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria;	Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 397)	
Matheis, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., and Hood, L.	

Sequence-tagged connectors: A sequence approach to mapping and

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahitab GG, Wallace JC, Hood L

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel.: (206) 616-3618
Tel.: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dong.med.buffalo.edu). Clones may be purchased from BACpac Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.hnsc.washington.edu>

High quality sequence stop: 397.
Location/Qualifiers

```

source
1. .397
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_plate=659 COL=9 Row=B"
/clone_id="RPCT-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
148 a 50 c 61 g 138 t

```

Query Match	42.9%;	Score 26.6;	DB 17;	Length 397;
Best Local Similarity	66.7%;	Pred. No. 4.be+02;		
Matches 38;	Conservative 0;	Mismatches 19;	Indels 0;	Gaps 0;

5	ATAGGTGGTGATTTGCGAAGACATAGAGGACACATCTAAAGTGGGAAATGTAA	61
DB	AGAGGTGGTGGAATTTTCAGAGAAAGATAGAGGAAACGTTAAATGAAAGATTA	372

RESULT_13	
LOCUS	A0620909/c
DEFINITION	453 bp DNA linear GSS 16-JUN-1999
ACCESSION	A0620909
VERSION	HS_2178_B1.D03.MF.CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2178 Col=5 Row=H, DNA sequence.
	A0620909.1 GI:5083389

ORGANISM
Homo sapiens
Eukaryotes: Metazoa; Chordata; Craniata; Vertebrata; Eucleleostomi;
Mammalia: Eutheria: Primates: Catarrhini; Homnidae; Homo.
1 (bases 1 to 453)
REFERENCE
Mahataas,G.G., Wallace,J.C., Smith,K., Swartzall,S., Holzman,T.,
Keller,A., Shakeri,R., Furlong,U., Young,U., Zhao,S., Adams,M.D., and
Hood,L.

TITLE	Sequence-tagged connectors: A sequence approach to mapping and
JOURNAL	examining the human genome
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
COMMENT	99380589
	Contact: Mahatras GG, Wallace JC, Hood L
	High Throughput Sequencing Center

401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@reagen.com).
BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 2178 row: H column: 5
Seq primer: T7
Clones: BAC ends
High quality sequence stop: 453.

FEATURES	source	Location/Qualifiers
		1. .453
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone="plate=2178 Col=5 Row=H"
		/clone_lib="CTR Approved Human Genomic Sperm Library D"
		/sex="male"
		/note="Organ: sperm; Vector: pBeloBac11; BAC clones in E-Coli DH10B"
BASE COUNT		141 a 59 c 79 g 169 t 5 others
ORIGIN		

Query Match	42.9%	Score 26.6	DB 17	Length 453
Best Local Similarity	65.5%	Pred. No. 5.1e+02		
Matches 38	Conservative 0	Mismatches 20	Indels 0	Gaps 0

5 ATTAGCTGGCGATTTCGAAAAGACACATAGAGACACATGCTTAAAGTGAATAATGTAAG 62

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 667.263 Seconds

(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-12

Perfect score: 63

Sequence: 1 ttccattatcaagcactgtg.....aacacactagaggtcccg 63

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_rts:*
12: gb_ry:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_rts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_hcg_hum:*
39: em_hcg_mus:*
40: em_hcg_hum:*
41: em_hcg_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	100.0	63	6	AX036746 Sequence
2	60	95.2	317	6	AX036737 Sequence
3	60	95.2	371	6	AX036739 Sequence
4	60	95.2	393	6	AX036753 Sequence
5	60	95.2	462	6	AX036754 Sequence
6	60	95.2	600	6	AX036756 Sequence
7	56	88.9	9285	6	AX039047 Sequence
8	56	88.9	15077	6	AX093052 Sequence
9	50.4	80.0	7489	14	CYMWCG
10	49	77.8	243	6	AX036735 Sequence
11	49	77.8	332	6	AX036755 Sequence
12	46	73.0	301	6	AX036741 Sequence
13	46	73.0	348	6	AX036738 Sequence
14	46	73.0	398	6	AX036740 Sequence
15	46	73.0	472	6	AX036759 Sequence
16	46	73.0	541	6	AX036758 Sequence
17	46	73.0	604	6	AX036757 Sequence
18	29.4	46.7	165061	2	AC107531 Rattus no
19	29	46.0	526	17	AG025881
20	29	46.0	148152	2	AC022200
21	29	46.0	153936	9	AP001207 Homo sapi
22	29	46.0	154587	9	AP001208 Homo sapi
23	28.8	45.7	206370	2	AL805937 Mus muscu
24	28	44.4	171777	2	AC073151 Mus muscu
25	28	44.4	215049	2	AC122807 Mus muscu
26	27.6	43.8	121027	2	AL157891 Homo sapi
27	27.6	43.8	128228	9	AL138810 Human DNA
28	27.6	43.8	128228	9	HS298105 Homo sapi
29	27.6	43.8	172039	2	AC099141 Rattus no
30	27.6	43.8	176425	2	AC090418 Homo sapi
31	27.6	43.8	221619	2	AC122844 Mus muscu
32	27.4	43.5	140741	9	HS29882
33	27.4	43.5	223678	10	AC027654 Mus muscu
34	27.2	43.2	181350	2	AC116233 Rattus no
35	27	42.9	174044	9	AC090675 Homo sapi
36	27	42.9	182261	2	AC026134 Homo sapi
37	27	42.9	256841	2	AC137284 Mus muscu
38	26.8	42.5	78770	2	AC034230 Homo sapi
39	26.8	42.5	112630	2	AC025461 Homo sapi
40	26.8	42.5	180789	2	AC024633 Homo sapi
41	26.8	42.5	194664	2	AC024627 Homo sapi
42	26.8	42.5	195369	9	AC113395 Homo sapi
43	26.8	42.5	213985	10	AL591129 Mouse DNA
44	26.6	42.2	55367	9	AC023905 Homo sapi
45	26.6	42.2	185771	2	AC111718 Rattus no

ALIGNMENTS

RESULT 1
AX036746 63 bp DNA
LOCUS AX036746
DEFINITION Sequence 12 from Patent WO0058485.
ACCESSION AX036746
VERSION AX036746.1 GI:11226255
KEYWORDS
ORGANISM
SOURCE
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 63)
Rance, I., Theisen, M. and Gruber, V.
AUTHORS Chimeric expression promoters originating from commelina yellow
TITLE Mottle virus and Cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 12 05-OCT-2000;

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
FEATURES
Source Location/Qualifiers
1..63
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional desoxynucleotide building block 55"
BASE COUNT 21 a 16 c 10 g 16 t
ORIGIN

Query Match 100.0%; Score 63; DB 6; Length 63;
Best Local Similarity 100.0%; Pred. No. 2.2e-10;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACAACAACCACTAGAGATCC 60
Db 1 TTCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACAACAACCACTAGAGATCC 60

Qy 61 CCG 63
Db 61 CCG 63

RESULT 2
AX036737 317 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 3 from Patent WO0058485.
ACCESSION AX036737
VERSION AX036737.1 GI:11226246
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
AUTHORS
1 (bases 1 to 317)
Rance,I., Theisen,M. and Gruber,V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 3 05-OCT-2000;
JOURNAL
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
Source Location/Qualifiers
1..317
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1116"
BASE COUNT 107 a 61 c 74 g 75 t
ORIGIN
promoter 1..317

Query Match 95.2%; Score 60; DB 6; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACAACAACCACTAGAGATCC 60
Db 258 TTCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACAACAACCACTAGAGATCC 317

RESULT 3
AX036739 371 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 5 from Patent WO0058485.
ACCESSION AX036739
VERSION AX036739.1 GI:11226248
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
AUTHORS
1 (bases 1 to 371)
Rance,I., Theisen,M. and Gruber,V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 5 05-OCT-2000;
JOURNAL
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
Source Location/Qualifiers

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
FEATURES
Source Location/Qualifiers
1..371
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1146"
BASE COUNT 122 a 68 c 89 g 92 t
ORIGIN
promoter 1..371

Query Match 95.2%; Score 60; DB 6; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACAACAACCACTAGAGATCC 60
Db 312 TTCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACAACAACCACTAGAGATCC 371

RESULT 4
AX036753 393 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 19 from Patent WO0058485.
ACCESSION AX036753
VERSION AX036753.1 GI:11226262
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
AUTHORS
1 (bases 1 to 393)
Rance,I., Theisen,M. and Gruber,V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 19 05-OCT-2000;
JOURNAL
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
Source Location/Qualifiers
1..393
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1162"
BASE COUNT 128 a 75 c 93 g 97 t
ORIGIN
promoter 1..393

Query Match 95.2%; Score 60; DB 6; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACAACAACCACTAGAGATCC 60
Db 334 TTCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACAACAACCACTAGAGATCC 393

RESULT 5
AX036754 462 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 20 from Patent WO0058485.
ACCESSION AX036754
VERSION AX036754.1 GI:11226263
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
AUTHORS
1 (bases 1 to 462)
Rance,I., Theisen,M. and Gruber,V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 20 05-OCT-2000;
JOURNAL
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
Source Location/Qualifiers

source 1. .462
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPRI163"

Promoter 1. .462
BASE COUNT 148 a 87 c 111 g 116 t
ORIGIN

Query Match 95.2%; Score 60; DB 6; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTATTAGACCTGTGTAGTACTTGAACCAACCAACCAACCACTTAGAGATCC 60
DB 403 TTCTTATTAGACCTGTGTAGTACTTGAACCAACCAACCAACCACTTAGAGATCC 462

RESULT 6
AX036756 AX036756 600 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 22 from Patent WO0058485.
ACCESSION AX036756
VERSION AX036756.1 GI:11226265
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 600)
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
PATENT: WO 0058485-A 22 05-OCT-2000;
JOURNAL MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
ORIGIN

FEATURES
source
1. .600
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MPRI165"

Promoter 1. .600
BASE COUNT 188 a 111 c 147 g 154 t
ORIGIN

Query Match 95.2%; Score 60; DB 6; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTATTAGACCTGTGTAGTACTTGAACCAACCAACCAACCACTTAGAGATCC 60
DB 541 TTCTTATTAGACCTGTGTAGTACTTGAACCAACCAACCAACCACTTAGAGATCC 600

RESULT 7
AX093047 AX093047 9285 bp DNA linear PAT 30-MAR-2001
LOCUS
DEFINITION Sequence 52 from Patent WO0118192.
ACCESSION AX093047
VERSION AX093047.1 GI:13509522
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 9285)
AUTHORS Gruber, V. and Comeau, D.
TITLE Synthetic vectors, transgenic plants containing them, and methods
for obtaining them
PATENT: WO 0118192-A 52 15-MAR-2001;
JOURNAL MERISTEM THERAPEUTICS (FR)
ORIGIN

FEATURES
source
1. .9285
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="PMRT1336"

misc_feature 1
/note="PMRT1336 results from the insertion into PMRT1196
of the promoter MPRI165 isolated from plasmid PMRT1322 as
described in PCR patent application PCT/IB00/00370"

Promoter 2440 a 2252 c 2506 g 2087 t
BASE COUNT
ORIGIN

Query Match 88.9%; Score 56; DB 6; Length 9285;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTATTAGACCTGTGTAGTACTTGAACCAACCAACCAACCACTTAGAGG 56
DB 6330 TTCTTATTAGACCTGTGTAGTACTTGAACCAACCAACCAACCACTTAGAGG 6385

RESULT 8
AX093052 AX093052 15077 bp DNA linear PAT 30-MAR-2001
LOCUS
DEFINITION Sequence 57 from Patent WO0118192.
ACCESSION AX093052
VERSION AX093052.1 GI:13509527
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE 1 (bases 1 to 15077)
AUTHORS Gruber, V. and Comeau, D.
TITLE Synthetic vectors, transgenic plants containing them, and methods
for obtaining them
PATENT: WO 0118192-A 57 15-MAR-2001;
JOURNAL MERISTEM THERAPEUTICS (FR)
ORIGIN

FEATURES
source
1. .15077
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="PMRT1342"

misc_feature 1
/note="PMRT1342 results from the replacement of the
expression cassette ep35-gus-polyA35S from PMRT1335 by
the expression cassette U5-gus-polyA35S isolated from
PMRT1336"

Promoter 3672 a 3892 c 4225 g 3288 t
BASE COUNT
ORIGIN

Query Match 88.9%; Score 56; DB 6; Length 15077;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTATTAGACCTGTGTAGTACTTGAACCAACCAACCAACCACTTAGAGG 56
DB 7369 TTCTTATTAGACCTGTGTAGTACTTGAACCAACCAACCAACCACTTAGAGG 7424

RESULT 9
CYMVG CYMVG 7489 bp DNA linear VRL 12-SEP-1993
LOCUS
DEFINITION Commelina yellow mottle virus complete circular genome.
ACCESSION X52938
VERSION X52938.1 GI:59047
KEYWORDS complete genome; protease; reverse transcriptase; ribonuclease H.
SOURCE Commelina yellow mottle virus.
ORGANISM Commelina yellow mottle virus
viruses; Retroviral viruses; Caulimoviridae; Badnavirus.
REFERENCE 1 (bases 1 to 7489)
AUTHORS Olszewski, N.E.
TITLE Direct Substitution
JOURNAL Submitted (02-MAY-1990) Olszewski N.E., University of Minnesota,
Dept. of Plant Biology, 220 Biological Sciences Center, 1445 Gortner
Ave., St. Paul, MN 55108, USA
(revised by [4])
REMARK 2 (bases 1 to 7489)
AUTHORS Medberry, S.L., Lockhart, B.E. and Olszewski, N.E.

TITLE Properties of Commelina yellow mottle virus's complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus

JOURNAL Nucleic Acids Res. 18 (18), 5505-5513 (1990)

MEDLINE 91016835

PUBMED 1699203

REFERENCE 3 (bases 1 to 7489)

AUTHORS Oleszewski, N.E.

TITLE Direct Submission

JOURNAL Submitted (29-NOV-1990) Oleszewski N.E., University of Minnesota, Dept of Plant Biology, 220 Biological Sciences Center, 1445 Gortner Ave., St. Paul, MN 55108, USA

REFERENCE 4 (bases 1 to 7489)

AUTHORS Oleszewski, N.

TITLE Direct Submission

JOURNAL Submitted (07-DEC-1990)

COMMENT [3] The extra c at position 2672 permits readthrough to result in one ORF (216 kd protein) rather than two (47 kd ORF and 163 kd ORF).

FEATURES

source 1..7489 location/Qualifiers

1..7489 /organism="Commelina yellow mottle virus"

/db_xref="taxon:10653"

/clone="pCOYMW89 (and pCOYMW89)"

1..23 /product="cRNA-Met"

/note="put. cRNA-Met (put. primer for minus strand synthesis)"

300..302 /note="gac was cga"

/citation=[1]

496..1098 /note="ORF 1"

/codon_start=1

/protein_id="CAA37108.1"

/db_xref="GI:59048"

/translation="MNWMLKSHPTPLGLPYSLDLPFCFMNOVDQVKLIDMLSSA KLSSEVIVFTPEVKIRLRLANIHILAHVLAIGFVIVLYVDIIFPLKTIQLOSO KESSENASTLVKQGRSLKQIEDLSYQSEBLATGRDYISRRPLSKQDELYV RISQPKFTKQTEALTEELKVEEYAKLISFKGNL"

541 /note="c in clone pCOYMW100"

873 /note="t in clone pCOYMW100"

1098..1505 /note="t in clone pCOYMW100"

/note="ORF 2"

/codon_start=1

/protein_id="CAA37109.1"

/db_xref="GI:59049"

/translation="MSNITKSGYKALSTNNYLAATIGGATDVGLTQKQNT TTVVAKITQIEDLSYQSEBLATGRDYISRRPLSKQDELYV RISQPKFTKQTEALTEELKVEEYAKLISFKGNL"

1338..1340 /note="3bp deletion in pCOYMW100"

1506..7166 /note="ORF 3, 216 kd protein"

/codon_start=1

/protein_id="CAA37110.1"

/db_xref="GI:59050"

/translation="MATRRUPAYTQDGTATSGVVEYEDQIRSRNDORRHITWA GRRRLISIMPGVSSERTLEMOMNPEVOQRSMNRBAEVPAAVLKRTFGSVNRV YSRSEKRMNVGSOVDKSFIOSSFEVLSRTGIEFTIHGWLVAHQILHRTFAGTM ALVFRDTRMSDDRAVLAAMEIDLSBNOQIVVLPIMMTIKSPYHIIQICWTAKCYD GWSGDNLLITRGLTGLSNTSNVGFAYDKAVAVHLOQSGVAVLIGEKWDARFANG OMNIERSKVVPWQPTEMKAVSNYDGTSLRPFENVYAASTSKPQVNEKDEEINBEO EINHSLNLIINDESDTDEDEYQOYRVAASGDSFTYDTGVAEIDRCNDLBEY VPSETSTPTDESEALIDDELAAYEORCSDSLOSQPRKRYETTPQSPPHLNE SRSSSSASTSMQDVEEIVRLMKERMKQKKKAQQAQQLSSQADEPTEENIENEN KQAOEPTOEIPTHKENOEPIQNEIHVEEPAPFKHLAQLSELVMAAESGQSG VGPQPVNAQPDVNMEGPAGYAPATSAQTSWNGVNIPIVKSANFRMKPGNQLPSAQ

GDGAMLVFGMNVSPFVPRMASITNYSISFNPDGDKIAMEDLLGTERKIFVS
WRMRFDYONIAKINODGGOALISQIRBIFLGDPTLQNTVONIAFRKIKOIVC
PYNQISIRYIMDYMTLAATGMLMSTBERPAISELFTGMPAIGSRVQAAYKIMPT
SAVNLPSRYFTTINYLTLEQCKASTKRSIALDPCDFTEGIGSGEKYTKARKA
TKTYGAHONNHLRYTKAKYQKCKCYTCQGBGHYANQCNKRDQORVAILOSILKE
NEEVVADKDEEBDEIFSVLGEDEYOEETIWLIEDDIOQIIKEFSKGDLSRRVVG
PNPFGPAEYQMGVLAKPKSSWRRPIQATLEENHNMVTAISTGLACRSCKOPLAVQ
CHHCAVYCFMCAEAYHDVQAEKILSKQVSPSARQKGAIVIEEDIEGEPILISLO
ONORLOKQVERLOEIMLREKXDSLSKRSASRSTESDOYSEALGINTIISYRG
FKEETRMATQLEKAITVDEBTEBERCMILBEKTEETISYTAQAEYNGLYNVKRG
IKPDNEPYYINAIIVDTGATACIIOISALPENYBAKATVRSVLGISTOMTKA
GRLIGEYFRMPVTVMMNGSLSPGIOMTIGSFISLGEARKKIDLEITFYGLVTSI
EISRTTOVANSIEIELESEDEYINIASVETSPFLQEFARKKIDLEKMEKRYIGE
NPMEFWKNNKIKCKLNIINPDIKIMGRPIKHYTPGDEAMTQOINLLQMKVIRPSES
KRSSTAFTVRSGETIDPITGCKKGRKRVFNYKLIINTEBDOYSEALGINTIISYRG
RSKITSKFLPKSGFMQVANAEBSPMTATLAKKLYELVMPFGLKNAAPAIQRKDN
VFKGTERTFLAVYIDILVSETAEQSHQLYMLQCKENGILSLTPKNIKIGPELIDF
LGSALGCTKIKIQPHIISKIDFSEDKLATPESGMSGLISLARYIODIGLVOPL
ROMKAPYGRMNPETRMVROIKKVKQLPDLOLPKDSFTIETDGCMTGGAVCX
WMSKHPDPRSTERICAVASGSPNPKSTIDAEIOAAIHGLDKXYLYLKKELIIRSD
CEALIKFYKTKTENKRSRVRVLPSPFLGIGITTYTPREIIDKHNGLADLSMINTPI
VERNDSPYRFTSSVADALVCNDHGRNLISAVINDITTLRR"

1970 /note="a in clone pCOYMW100"

2452 /note="t in clone pCOYMW100"

2672..2674 /note="ccc was cc in [1]"

/citation=[1]

2977 /note="c in clone pCOYMW100"

4611 /note="a in clone pCOYMW100"

4794 /note="t in clone pCOYMW100"

6302 /note="t in clone pCOYMW100"

6434 /note="c in clone pCOYMW100"

7378 /note="c in clone pCOYMW100"

7462 /note="c in clone pCOYMW100"

7478 /note="additional t in pCOYMW100"

7482 /note="c in clone pCOYMW100"

/note="g in clone pCOYMW100"

BASE COUNT 2662 a 1389 c 1577 g 1861 t

ORIGIN

Query Match 80.0%; Score 50.4; DB 14; Length 7489;
Best Local Similarity 90.0%; Pred. No. 1.5e-06;
Matches 54; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TTCCTATTACACCTGTGTAGTACTAGAAACCAACACACCTAGAGATCC 60
|||||
Db 7318 TTCCTATTACACCTGTGTAGTACTAGAAACCAACACACCTAGAGATAC 7377

RESULT 10

AX036735 243 bp DNA linear PART 16-NOV-2000

LOCUS AX036735

DEFINITION Sequence 1 from Patent WO0058485.

ACCESSION AX036735

VERSION AX036735.1 GI:11226244

KEYWORDS

ORGANISM synthetic construct.

SOURCE artificial sequences.

REFERENCE 1 (bases 1 to 243)

AUTHORS Range, I., Theisen, M. and Gruber, V.

TITLE Chimeric expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus

JOURNAL Patent: WO 0058485-A 1 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. .243
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="243 bp promoter fragment from the intergenic region
of Commelina Yellow Mottle Virus"
promoter
1. .243
BASE COUNT 71 a 53 c 45 g 74 t
ORIGIN
Query Match 77.8%; Score 49; DB 6; Length 243;
Best Local Similarity 100.0%; Pred. No. 7.5e-06;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTTATTAGCACTGTGTAGTACCTTGAACCAACCAACACAC 49
Db 195 TTCCTTATTAGCACTGTGTAGTACCTTGAACCAACCAACACAC 243
RESULT 11
AX036755 392 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 21 from Patent WO0058485.
ACCESSION AX036755
VERSION AX036755.1 GI:11226264
KEYWORDS
SOURCE
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 392)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 21 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. .392
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1164"
promoter
1. .392
BASE COUNT 127 a 80 c 87 g 98 t
ORIGIN
Query Match 77.8%; Score 49; DB 6; Length 392;
Best Local Similarity 100.0%; Pred. No. 6.9e-06;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTTATTAGCACTGTGTAGTACCTTGAACCAACCAACACAC 49
Db 334 TTCCTTATTAGCACTGTGTAGTACCTTGAACCAACCAACACAC 382
RESULT 12
AX036741 301 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 7 from Patent WO0058485.
ACCESSION AX036741
VERSION AX036741.1 GI:11226250
KEYWORDS
SOURCE
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 301)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 7 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)

FEATURES ; GRUBER VERONIQUE (FR)
source
1. .301
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1154"
promoter
1. .301
BASE COUNT 98 a 54 c 74 g 75 t
ORIGIN
Query Match 73.0%; Score 46; DB 6; Length 301;
Best Local Similarity 100.0%; Pred. No. 7.1e-05;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 ACTGTGTAGTACCTTGAACCAACCAACCAACCTAGAGATCC 60
Db 256 ACTGTGTAGTACCTTGAACCAACCAACCAACCTAGAGATCC 301
RESULT 13
AX036738 348 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 4 from Patent WO0058485.
ACCESSION AX036738
VERSION AX036738.1 GI:11226247
KEYWORDS
SOURCE
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 348)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 4 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. .348
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter Mp1117"
promoter
1. .348
BASE COUNT 116 a 70 c 78 g 84 t
ORIGIN
Query Match 73.0%; Score 46; DB 6; Length 348;
Best Local Similarity 100.0%; Pred. No. 6.9e-05;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 ACTGTGTAGTACCTTGAACCAACCAACCAACCTAGAGATCC 60
Db 303 ACTGTGTAGTACCTTGAACCAACCAACCAACCTAGAGATCC 348
RESULT 14
AX036740 398 bp DNA linear PAT 16-NOV-2000
LOCUS
DEFINITION Sequence 6 from Patent WO0058485.
ACCESSION AX036740
VERSION AX036740.1 GI:11226249
KEYWORDS
SOURCE
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 398)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
TITLE
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
JOURNAL Patent: WO 0058485-A 6 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
FEATURES
source
1. .398
Location/Qualifiers

/organism="synthetic construct"
/db xref="taxon:32630"
/note="promoter MPr1147"
1..398
BASE COUNT 128 a 80 c 93 g 97 t
ORIGIN

Query Match 73.0%; Score 46; DB 6; Length 398;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 ACTTGTTAGTAGCTTAGAAAAACCAACAACTAGAGATCC 60
DB 353 ACTTGTTAGTAGCTTAGAAAAACCAACAACTAGAGATCC 398

RESULT 15
AX036759 472 bp DNA linear PAT 16-NOV-2000
LOCUS AX036759
DEFINITION Sequence 25 from Patent WO0058485.
ACCESSION AX036759
VERSION AX036759.1 GI:11226268
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 472)

REFERENCE
AUTHORS Rance, I., Theisen, M. and Gruber, V.
TITLE ChimERIC expression promoters originating from commelina yellow
JOURNAL mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 25 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
location/Qualifiers

FEATURES
source 1..472
/organism="synthetic construct"
/db xref="taxon:32630"
/note="promoter MPr1169"
1..472
BASE COUNT 149 a 92 c 112 g 119 t
ORIGIN
promoter

Query Match 73.0%; Score 46; DB 6; Length 472;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 ACTTGTTAGTAGCTTAGAAAAACCAACAACTAGAGATCC 60
DB 427 ACTTGTTAGTAGCTTAGAAAAACCAACAACTAGAGATCC 472

Search completed: May 11, 2003, 06:04:58
Job time : 672.263 secB

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 122.526 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-12

Perfect score: 63
Sequence: 1 ttccctatcctaagcactgt.....aacacactagagatccccc 63

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

1: N_Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	100.0	63	21	AAA96846
2	60	95.2	317	21	AAA96837
3	60	95.2	371	21	AAA96839
4	60	95.2	393	21	AAA96853
5	60	95.2	462	21	AAA96854
6	60	95.2	600	21	AAA96856
7	49	77.8	243	21	AAA96835
8	49	77.8	392	21	AAA96855
9	46	73.0	301	21	AAA96841

10	46	73.0	348	21	AAA96838	Nucleotide sequenc
11	46	73.0	398	21	AAA96840	Nucleotide sequenc
12	46	73.0	472	21	AAA96859	Nucleotide sequenc
13	46	73.0	541	21	AAA96858	Nucleotide sequenc
14	46	73.0	604	21	AAA96857	Nucleotide sequenc
15	27	42.9	2928	23	AA589944	DNA encoding novel
16	27	42.9	4600	23	AA568229	DNA encoding novel
17	26.4	41.9	2786	22	AAH14506	Human cDNA sequenc
18	26	41.3	4289	23	ABP03686	Drosophila melanog
19	25.8	41.0	14919	22	AA546505	Tumour suppressor
20	25	39.7	694	21	AA14716	Aspergillus oryzae
21	24.8	39.4	815	21	AA15867	Human prostate can
22	24.8	39.4	1969	22	AAH14864	Human cDNA sequenc
23	24.6	39.0	1196	24	ABL40669	Human Fe-S protein
24	24.4	38.7	2203	22	AA160067	Human polynucleoti
25	24.4	38.7	2248	22	AA158281	Human polynucleoti
26	24.4	38.7	2915	22	AA521348	Human cDNA sequenc
27	24.4	38.7	52562	22	AAK86669	Human immune/haema
28	24.4	38.7	53075	22	AAK86671	Human immune/haema
29	24	38.1	329	22	AA187625	Human polynucleoti
30	24	38.1	777	24	AAH33092	Streptococcus poly
31	24	38.1	1023	22	AAH36600	Human colon cancer
32	24	38.1	1032	20	AAZ27434	Human secreted pro
33	24	38.1	1034	20	AAZ27315	Human secreted pro
34	24	38.1	3245	22	AAH54626	S. epidermidis gen
35	23.8	37.8	10202	23	ABP02571	Drosophila melanog
36	23.8	37.8	10369	22	AA546304	Tumour suppressor
37	23.8	37.8	10369	24	ABU32393	Human immune syste
38	23.8	37.8	19603	23	ABU02570	Drosophila melanog
39	23.6	37.5	1203	23	AA580397	DNA encoding novel
40	23.6	37.5	2462	23	ABU22798	Drosophila melanog
41	23.6	37.5	5102	24	ABU92274	Chemically treated
42	23.4	37.1	6847	23	ABU14139	Drosophila melanog
43	23.4	37.1	14872	19	AAV52205	Streptococcus pneu
44	23.4	37.1	22948	23	ABU14138	Drosophila melanog
45	23.2	36.8	411	23	ABV46444	Human prostate exp

ALIGNMENTS

RESULT 1	
AAA96846	AAA96846 standard; DNA; 63 BP.
19-FEB-2001 (first entry)	
Directional deoxynucleotide building block 55.	
Promoter; intergenic region; Commelina yellow mottle virus;	
chimeric expression promoter; plant vascular expression promoter;	
plant green tissue expression promoter; Cassava vein mosaic virus;	
transgenic plant; ss.	
Synthetic.	
WO200058485-A1.	
05-OCT-2000.	
29-MAR-2000; 2000MO-IB00370.	
29-MAR-1999; 99FR-0003925.	
(MERI-) MERISTEM THERAPEUTICS.	
Rance I, Gruber V, Theisen M;	
WPI; 2000-647238/62.	
Chimeric expression promoter for transgenic plant production, comprises	

The present sequence represents a chimeric promoter of the invention. The specification describes chimeric expression promoters. These chimeric promoters comprise a nucleic acid sequence which is derived from a first plant promoter, in which a plant vascular expression promoter region is replaced with a nucleic acid sequence derived from a second plant promoter comprising a plant green tissue expression promoter region. Preferably, the first plant promoter originates from Commelina yellow mottle virus, the first plant promoter originates from the Cassava vein mosaic virus, and the second plant promoter originates from the Cassava vein mosaic virus. Especially, the promoters are derived from intergenic regions. The chimeric promoters are useful for producing transgenic plants.

Sequence 317 BP; 107 A; 61 C; 74 G; 75 T; 0 other;

Query Match 95.2%; Score 60; DB 21; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TTCTTATTATAGACCTGCTAGACTAGAACAACCAACAACTAGAGATCC 60
Db 258 TTCCTATTATAGACTGTGTAGCTTAGAAAACCAACAACTAGAGATCC 317

RESULT 3
AAA96839
ID AAA96839 standard; DNA; 371 BP.
XX
AC AAA96839;
XX
AT 19-FEB-2001 (first entry)
DE Nucleotide sequence of chimeric expression promoter MP1146.
XX
KW Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KN plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN W0200058485-A1.
XX
PD 05-OCT-2000.
XX
PP 29-MAR-2000; 2000MO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WP1; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
PS Claim 5; Page 81; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
XX The specification describes chimeric expression promoters. These
XX chimeric promoters comprise a nucleic acid sequence which is derived
XX from a first plant promoter, in which a plant vascular expression
XX promoter region is replaced with a nucleic acid sequence derived from
XX a second plant promoter comprising a plant green tissue expression
XX promoter region. Preferably, the first plant promoter originates from
XX Commelina yellow mottle virus, and the second plant promoter originates
XX from the Cassava vein mosaic virus. Especially, the promoters are
XX derived from intergenic regions. The chimeric promoters are useful
XX for producing transgenic plants.

XX SQ Sequence 371 BP; 122 A; 68 C; 89 G; 92 T; 0 other;
Query Match 95.2%; Score 60; DB 21; Length 371;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTTATTAGACCTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 60
DB 312 TTCCTTATTAGACCTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 371
RESULT 4
ID AAA96853
AAA96853 standard; DNA; 393 BP.
AC AAA96853;
XX
XX 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1162.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
XX transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
XX
XX 29-MAR-2000; 2000MO-IB00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
XX Claim 5; Page 85; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
XX Sequence 393 BP; 128 A; 75 C; 93 G; 97 T; 0 other;
SO
Query Match 95.2%; Score 60; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 1.1e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTTATTAGACCTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 60
DB 334 TTCCTTATTAGACCTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 393

RESULT 5
ID AAA96854
AAA96854 standard; DNA; 462 BP.
XX
XX AAA96854;
XX
XX 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1163.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
XX transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
XX
XX 29-MAR-2000; 2000MO-IB00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
XX Claim 5; Page 86; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
XX Sequence 462 BP; 148 A; 87 C; 111 G; 116 T; 0 other;
SQ
Query Match 95.2%; Score 60; DB 21; Length 462;
Best Local Similarity 100.0%; Pred. No. 1.2e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTTATTAGACCTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 60
DB 403 TTCCTTATTAGACCTTGTTAGTACTTAGAAACCAACACAACTAGAGATCC 462
RESULT 6
ID AAA96856
AAA96856 standard; DNA; 600 BP.
XX
XX AAA96856;
XX
XX 19-FEB-2001 (first entry)
XX
XX Nucleotide sequence of chimeric expression promoter MP1165.
XX
XX

KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Claim 5; Page 86-87; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 600 BP; 188 A; 111 C; 147 G; 154 T; 0 other;
Query Match 95.2%; Score 60; DB 21; Length 600;
Best Local Similarity 100.0%; Pred. No. 1,2e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTAAAGCAGCTGTGTAGTACTTGAAGAAACCAACCAACCTAGAGATCC 60
DB 541 TTCCTATTAAAGCAGCTGTGTAGTACTTGAAGAAACCAACCAACCTAGAGATCC 600
RESULT 7
AAA96835
ID AAA96835 standard; DNA; 243 BP.
XX
AC AAA96835;
XX
DT 19-FEB-2001 (first entry)
XX
DE Promoter from intergenic region of Commelina yellow mottle virus.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; ss.
XX
OS Commelina yellow mottle virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX

PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Claim 4; Page 79; 91pp; English.
XX
XX The present sequence represents a promoter fragment from the intergenic
CC region of Commelina yellow mottle virus. The promoter is a strong
CC promoter in vascular and reproductive tissues. The promoter is used to
CC construct chimeric expression promoters. These chimeric promoters
CC comprise a nucleic acid sequence which is derived from a first
CC plant promoter, in which a plant vascular expression promoter region is
CC replaced with a nucleic acid sequence derived from a second plant
CC promoter comprising a plant green tissue expression promoter region.
CC Preferably, the first plant promoter originates from Commelina yellow
CC mottle virus, and the second plant promoter originates from the Cassava
CC vein mosaic virus. The chimeric promoters are useful for producing
CC transgenic plants.
XX
SQ Sequence 243 BP; 71 A; 53 C; 45 G; 74 T; 0 other;
Query Match 77.8%; Score 49; DB 21; Length 243;
Best Local Similarity 100.0%; Pred. No. 7.7e-08;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTCCTATTAAAGCAGCTGTGTAGTACTTGAAGAAACCAACCAACCAAC 49
DB 195 TTCCTATTAAAGCAGCTGTGTAGTACTTGAAGAAACCAACCAACCAAC 243
RESULT 8
AAA96855
ID AAA96855 standard; DNA; 392 BP.
XX
AC AAA96855;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter Mpr164.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX

PT Chimeric expression promoter for transgenic plant production, comprising
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
PS Claim 5; Page 86; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 392 BP, 127 A; 80 C; 87 G; 98 T; 0 other;
XX
Query Match 77.8%; Score 49; DB 21; Length 392;
Best Local Similarity 100.0%; Pred. No. 8,4e-08;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps
XX
Qy 1 TTCCTTATTTAAGCACTTGTGTAGTACCTTAGAAAACCAACACACAC 49
Db 334 TTCCTTATTTAAGCACTTGTGTAGTACCTTAGAAAACCAACACACAC 382
XX
RESULT 9
ID AAA96841 standard; DNA; 301 BP.
XX
XX AAA96841;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1154.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
XX chimeric expression promoter; plant vascular expression promoter;
XX plant green tissue expression promoter; Cassava vein mosaic virus;
XX transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
OS
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
PD
PD 29-MAR-2000; 2000WO-1B00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
PA
XX
XX Rance I, Gruber V, Theisen M;
PI
PI WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprising
XX sequence from promoter comprising vascular expression region replaced
XX with sequence from promoter comprising green tissue expression region
XX
XX Claim 5; Page 82; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
XX The specification describes chimeric expression promoters. These
XX chimeric promoters comprise a nucleic acid sequence which is derived
XX from a first plant promoter, in which a plant vascular expression
XX promoter region is replaced with a nucleic acid sequence derived from
XX a second plant promoter comprising a plant green tissue expression
XX promoter region. Preferably, the first plant promoter originates from
XX Commelina yellow mottle virus, and the second plant promoter originates
XX from the Cassava vein mosaic virus. Especially, the promoters are
XX derived from intergenic regions. The chimeric promoters are useful
XX for producing transgenic plants.

CC promote region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

XX Sequence 301 BP; 98 A; 54 C; 74 G; 75 T; 0 other;

50

Query Match 73.0%; Score 46; DB 21; Length 301;
Best Local Similarity 100.0%; Pred. No. 9.2e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps

0y 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCAACTAGAGATCC 60
Db 256 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCAACTAGAGATCC 301

RESULT 10
ID AAA96838
AC AAA96838 standard; DNA; 348 BP.
XX AAA96838;
AC
XX 19-FEB-2001 (first entry)
DT
XX
XX Nucleotide sequence of chimeric expression promoter MP1117.
DE
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
XX
OS Chimeric - Commelina yellow mottle virus.
OS
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO20058485-A1.
XX
XX 05-OCT-2000.
PD
XX
XX 29-MAR-2000; 2000MO-IB00370.
PP
XX
XX 29-MAR-1999; 99FR-0003925.
PR
XX
XX (MERI-) MERISTEM THERAPEUTICS.
PA
XX
XX Rance I, Gruber V, Theisen M;
PI
XX
XX WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
XX
XX Claim 5; Page 81; 91pp; English.

PS The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.

80 Sequence 348 BP; 116 A; 70 C; 78 G; 84 T; 0 other;

73.0%; Score 46; DB 21; Length 348;
Query Match

Best Local Similarity 100.0%; Pred. No. 9,4e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 ACTGTGTAGTAGCTTAAACCAACCAACCTTAGAGATCC 60
DB 303 ACTGTGTAGTAGCTTAAACCAACCAACCACTTAGAGATCC 348

RESULT 11

AAA96840
ID AAA96840 standard; DNA; 398 BP.

AC AAA96840;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1147.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KW chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

PD 05-OCT-2000.

PF 29-MAR-2000; 2000WO-1B00370.

PR 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

PI Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 82; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

CC The specification describes chimeric expression promoters. These

CC chimeric promoters comprise a nucleic acid sequence which is derived

CC from a first plant promoter, in which a plant vascular expression

CC promoter region is replaced with a nucleic acid sequence derived from

CC a second plant promoter comprising a plant green tissue expression

CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates

CC from the Cassava vein mosaic virus. Especially, the promoters are

CC derived from intergenic regions. The chimeric promoters are useful

CC for producing transgenic plants.

SQ Sequence 398 BP; 128 A; 80 C; 93 G; 97 T; 0 other;

Query Match 73.0%; Score 46; DB 21; Length 398;

Best Local Similarity 100.0%; Pred. No. 9.7e-07;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 ACTGTGTAGTAGCTTAAACCAACCAACCTTAGAGATCC 60

DB 353 ACTGTGTAGTAGCTTAAACCAACCAACCACTTAGAGATCC 398

RESULT 12

AAA96859
ID AAA96859 standard; DNA; 472 BP.

XX AAA96859;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1169.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KW chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

OS Chimeric - Commelina yellow mottle virus.

OS Chimeric - Cassava vein mosaic virus.

PN WO200058485-A1.

PD 05-OCT-2000.

PF 29-MAR-2000; 2000WO-1B00370.

PR 29-MAR-1999; 99FR-0003925.

PA (MERI-) MERISTEM THERAPEUTICS.

PI Rance I, Gruber V, Theisen M;

DR WPI; 2000-647238/62.

XX Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced

PT with sequence from promoter comprising green tissue expression region

PS Claim 5; Page 88; 91pp; English.

XX The present sequence represents a chimeric promoter of the invention.

CC The specification describes chimeric expression promoters. These

CC chimeric promoters comprise a nucleic acid sequence which is derived

CC from a first plant promoter, in which a plant vascular expression

CC promoter region is replaced with a nucleic acid sequence derived from

CC a second plant promoter comprising a plant green tissue expression

CC promoter region. Preferably, the first plant promoter originates from

CC Commelina yellow mottle virus, and the second plant promoter originates

CC from the Cassava vein mosaic virus. Especially, the promoters are

CC derived from intergenic regions. The chimeric promoters are useful

CC for producing transgenic plants.

SQ Sequence 472 BP; 149 A; 92 C; 112 G; 119 T; 0 other;

Query Match 73.0%; Score 46; DB 21; Length 472;

Best Local Similarity 100.0%; Pred. No. 1e-06;

Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 ACTGTGTAGTAGCTTAAACCAACCAACCACTTAGAGATCC 60

DB 427 ACTGTGTAGTAGCTTAAACCAACCAACCAACCTTAGAGATCC 472

RESULT 13

AAA96858
ID AAA96858 standard; DNA; 541 BP.

AC AAA96858;

DT 19-FEB-2001 (first entry)

DE Nucleotide sequence of chimeric expression promoter MP1168.

XX Promoter; intergenic region; Commelina yellow mottle virus;

KW chimeric expression promoter; plant vascular expression promoter;

KW plant green tissue expression promoter; Cassava vein mosaic virus;

XX transgenic plant; chimera; ss.

XX OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
XX PD
XX PF 29-MAR-2000; 2000WO-IB00370.
XX PR 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX PA
XX PI Rance I, Gruber V, Theisen M;
XX DR WPI; 2000-647238/62.
XX
XX PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
XX Claim 5; Page 87-88; 91pp; English.
XX PS
XX CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
XX Sequence 541 BP; 169 A; 104 C; 130 G; 138 T; 0 other;
SQ
XX
XX Query Match 73.0%; Score 46; DB 21; Length 541;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 ACTGTGTAGTACTGAGAAACCAACCAACCTAGAGATCC 60
Db 496 ACTGTGTAGTACTGAGAAACCAACCAACCTAGAGATCC 541
XX
XX RESULT 14
XX ID AAA96857
XX ID AAA96857 standard; DNA; 604 BP.
XX AC
XX AAA96857;
XX
XX 19-FEB-2001 (first entry)
XX DT
XX
XX Nucleotide sequence of chimeric expression promoter MPr1167.
XX DE
XX KW Promoter; intergenic region; Commelina yellow mottle virus;
KW chimeric expression promoter; plant vascular expression promoter;
KW plant green tissue expression promoter; Cassava vein mosaic virus;
KW transgenic plant; chimera; ss.
XX
XX OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX PN
XX PD 05-OCT-2000.
XX PD
XX PF 29-MAR-2000; 2000WO-IB00370.
XX PF
XX PR 29-MAR-1999; 99FR-0003925;
XX PR

XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX PA
XX PI Rance I, Gruber V, Theisen M;
XX DR WPI; 2000-647238/62.
XX
XX PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PT
XX
XX Claim 5; Page 87; 91pp; English.
XX PS
XX CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
XX Sequence 604 BP; 186 A; 116 C; 145 G; 157 T; 0 other;
SQ
XX
XX Query Match 73.0%; Score 46; DB 21; Length 604;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 ACTGTGTAGTACTGAGAAACCAACCAACCTAGAGATCC 60
Db 559 ACTGTGTAGTACTGAGAAACCAACCAACCTAGAGATCC 604
XX
XX RESULT 15
XX ID AAS89944
XX ID AAS89944 standard; cDNA; 2928 BP.
XX AC
XX AAS89944;
XX
XX 13-FEB-2002 (first entry)
XX DT
XX
XX DNA encoding novel human diagnostic protein #25748.
XX DE
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX OS
XX WO200175067-A2.
XX PN
XX PD 11-OCT-2001.
XX PD
XX PF 30-MAR-2001; 2001WO-US08631.
XX PF
XX 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PR
XX (HYSE-) HYSEQ INC.
XX PA
XX PI Dmanac RT, Liu C, Tang YT;
XX PI
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG25757.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX

PS Claim 1; SEQ ID No 25748; 103bp; English.

XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcc_sequences.

XX
SQ Sequence 2928 BP; 742 A; 773 C; 737 G; 676 T; 0 other;

Query Match 42.9%; Score 27; DB 23; Length 2928;

Best Local Similarity 85.7%; Pred. No. 7;
Matches 30; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 CTTAGAAAACCAACACCACTTAGAGATCCCC 62

DB 903 CTTGCTAGCCACACACAGCTTAGAGATCCCC 937

Search completed: May 11, 2003, 03:04:03
Job time : 124.526 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 25.2632 Seconds
(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-12

Perfect score: 63
Sequence: 1 ttccattatcaactgtg.....aacacactagagatcccg 63

Scoring table: IDENTITY_NOC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	24	38.1	1032	US-09-227-357-134	Sequence 134, App
2	24	38.1	1034	US-09-227-357-15	Sequence 15, Appl
3	23.4	37.1	14872	US-08-961-527-72	Sequence 37, Appl
4	22.4	35.6	518	US-09-280-116-245	Sequence 245, App
5	22.2	35.2	836	US-08-911-434A-5	Sequence 5, Appl
6	22	34.9	2885	US-08-920-812-4	Sequence 4, Appl
7	22	34.9	2885	US-08-920-827-4	Sequence 4, Appl
8	22	34.9	2885	US-08-921-177-4	Sequence 4, Appl
9	22	34.9	2885	US-08-362-577C-4	Sequence 4, Appl
10	22	34.9	2885	US-08-920-828-4	Sequence 4, Appl
11	21.8	34.6	96	US-08-755-587-4	Sequence 4, Appl
12	21.8	34.6	1056	US-08-755-587-28	Sequence 28, Appl
13	21.8	34.6	1917	US-08-755-587-1	Sequence 1, Appl
14	21.8	34.6	3131	US-08-840-466A-20	Sequence 20, Appl
15	21.8	34.6	3131	US-09-696-188B-20	Sequence 20, Appl
16	21.8	34.6	7240	US-08-755-587-15	Sequence 15, Appl
17	21.8	34.6	11283	US-08-603-753D-3	Sequence 3, Appl
18	21.8	34.6	11283	US-09-099-753-3	Sequence 3, Appl
19	21.8	34.6	11283	US-08-966-106-3	Sequence 3, Appl
20	21.8	34.6	11385	US-08-639-501-1	Sequence 1, Appl
21	21.8	34.6	11385	US-09-044-946-1	Sequence 1, Appl
22	21.8	34.6	11385	US-09-044-908-1	Sequence 1, Appl
23	21.8	34.6	19718	US-08-961-527-99	Sequence 99, Appl
24	21.8	34.6	43360	US-09-453-702B-206	Sequence 206, App
25	21.6	34.6	45325	US-09-453-702B-261	Sequence 261, App
26	21.6	34.3	444	US-07-708-038-3	Sequence 3, Appl
27	21.6	34.3	444	US-08-127-995-3	Sequence 3, Appl

28	21.6	34.3	1900	US-08-764-343-2	Sequence 2, Appl
C 29	21.6	34.3	1960	US-09-513-057C-12	Sequence 12, Appl
C 30	21.6	34.3	3617	US-09-513-057C-14	Sequence 14, Appl
C 31	21.4	34.0	272	US-09-537-357-31	Sequence 31, Appl
C 32	21.4	34.0	1715	US-09-537-357-56	Sequence 56, Appl
C 33	21.4	34.0	4471	US-08-615-942A-1	Sequence 1, Appl
C 34	21.2	33.7	198	US-09-134-001C-309	Sequence 309, App
C 35	21.2	33.7	7432	US-07-852-260-1	Sequence 1, Appl
C 36	21.2	33.7	7432	US-08-461-503-1	Sequence 1, Appl
C 37	21.2	33.7	7432	US-08-465-250-1	Sequence 1, Appl
C 38	21.2	33.7	11721	US-09-026-039-3	Sequence 3, Appl
C 39	21.2	33.7	11721	US-09-026-039-3	Sequence 3, Appl
C 40	21.2	33.7	112132	US-09-741-150-3	Sequence 3, Appl
C 41	21	33.3	375	US-09-134-001C-2507	Sequence 2507, App
C 42	21	33.3	1029	US-09-134-001C-1353	Sequence 1353, App
C 43	21	33.3	1993	US-08-487-890A-108	Sequence 108, App
C 44	21	33.3	1993	US-08-478-435-108	Sequence 108, App
C 45	21	33.3	1993	US-08-337-483-108	Sequence 108, App

ALIGNMENTS

RESULT 1
US-09-227-357-134
; Sequence 134, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13664
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18

```

1  EARLIER APPLICATION NUMBER: 60/055,949
2  EARLIER FILING DATE: 1997-08-18
3  EARLIER APPLICATION NUMBER: 60/055,953
4  EARLIER FILING DATE: 1997-08-18
5  EARLIER APPLICATION NUMBER: 60/055,950
6  EARLIER FILING DATE: 1997-08-18
7  EARLIER APPLICATION NUMBER: 60/055,947
8  EARLIER FILING DATE: 1997-08-18
9  EARLIER APPLICATION NUMBER: 60/055,964
10 EARLIER FILING DATE: 1997-08-18
11 EARLIER APPLICATION NUMBER: 60/056,360
12 EARLIER FILING DATE: 1997-08-18
13 EARLIER APPLICATION NUMBER: 60/055,684
14 EARLIER FILING DATE: 1997-08-18
15 EARLIER APPLICATION NUMBER: 60/055,984
16 EARLIER FILING DATE: 1997-08-18
17 EARLIER APPLICATION NUMBER: 60/055,954
18 EARLIER FILING DATE: 1997-08-18
19 EARLIER APPLICATION NUMBER: 60/058,785
20 EARLIER FILING DATE: 1997-09-12
21 EARLIER APPLICATION NUMBER: 60/058,664
22 EARLIER FILING DATE: 1997-09-12
23 EARLIER APPLICATION NUMBER: 60/058,660
24 EARLIER FILING DATE: 1997-09-12
25 EARLIER APPLICATION NUMBER: 60/058,661
26 EARLIER FILING DATE: 1997-09-12
27 NUMBER OF SEQ ID NOS: 672
28 SOFTWARE: PatentIn Ver. 2.0
29 SEQ ID NO 134
30 LENGTH: 1032
31 TYPE: DNA
32 ORGANISM: Homo sapiens
33 FEATURE:
34 NAME/KEY: SITE
35 LOCATION: (5)
36 OTHER INFORMATION: n equals a,t,g, or c
37 FEATURE:
38 NAME/KEY: SITE
39 LOCATION: (593)
40 OTHER INFORMATION: n equals a,t,g, or c
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
91
```

Oy	1	TTCCTATTTAAGCACTGTGTAGTAGCTTACGAAACCAACACACAA	48
Db	983	TTCAATTATTAAAGCACTGGTTCATTCACATGAAAAAAAAAAAAAAAA	1030
		Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;	
		Best Local Similarity 68.8%; Score 24; DB 4; Length 1032;	
		Query Match 38.1%; Pred. No. 8;	

RESULT 27-357-15
US-09-227-357-15
Sequence 15, Application US/09227357
Patent No. 6342581
GENERAL INFORMATION:
APPLICANT: Fischer et al.
TITLE OF INVENTION: 123 Human Secreted Proteins
FILE REFERENCE: P201091
CURRENT APPLICATION NUMBER: US/09/227,357
CURRENT FILING DATE: 1999-01-08
EARLIER APPLICATION NUMBER: PCT/US98/13684
EARLIER FILING DATE: 1998-07-07
EARLIER APPLICATION NUMBER: 60/051,926
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,793
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,925
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,929
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,803
EARLIER FILING DATE: 1997-07-08

```

EARLIER APPLICATION NUMBER: 60/052,732
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,931
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,932
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,916
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,930
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,918
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,920
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,733
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/052,795
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,919
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/051,928
EARLIER FILING DATE: 1997-07-08
EARLIER APPLICATION NUMBER: 60/055,722
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,723
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,948
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 1034
TYPE: DNA
ORGANISM: Homo sapiens
US-09-227-357-15

Query Match      38.1%; Score 24; DB 4; Length 1034;
Best Local Similarity 68.8%; Pred. No. 8;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

Qy 1 TTCCTATTAAAGCACTGTGTAAGCTTAGAAAAACCAACAACA 48
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 980 TTCATTATTTAAGCACTGGTTCATTCACTGAATAAAAAAAAAA 1027

RESULT 3
US-08-961-527-72/c


```
Sequence 72, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 14872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-72

Query Match          37.1%; Score 23.4; DB 4; Length 14872;
Best Local Similarity 63.2%; Pred. No. 22;
Matches 36; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1 TTCCTATTATGAGCTGTGATGCTTAGAAACCAACCAACCACTAGAGGA 57
Db 1596 TCCTGCAITTAACATATGATAGTACCAAGATACCGCAAAACAGCTAGAGGA 1540

RESULT 4
US-09-280-116-245
Sequence 245, Application US/09280116A
Patent No. 6331427
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
FILE REFERENCE: 5800-24, 035800/176965
CURRENT APPLICATION NUMBER: US/09/280,116A
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 268
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 245
LENGTH: 518
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
US-09-280-116-245

Query Match          35.6%; Score 22.4; DB 4; Length 518;
Best Local Similarity 62.5%; Pred. No. 26;
Matches 35; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

```
QY 7 ATTATGACCTGTGTAGTAGCTTAGAAACCAACCAACCACTAGAGATCCCC 62
Db 313 ATTATATCGACGATTTGGAGAGACGAGAAAATTTCAACACAAATTTGATTTCCCC 368

RESULT 5
US-08-911-434A-5/c
Sequence 5, Application US/08911434A
Patent No. 5959176
GENERAL INFORMATION:
APPLICANT: TORIKAI, Satomi
APPLICANT: OEDA, Kenji
TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: P.O. BOX 747
CITY: FALLS CHURCH
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22040
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,434A
FILING DATE: 12-AUG-1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Raymond C.
REGISTRATION NUMBER: 21,066
REFERENCE/DOCKET NUMBER: 2185-0199P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)205-8000
TELEFAX: (703)205-8050
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 836 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Daucus carota L.
INDIVIDUAL ISOLATE: Kuroda Gosun
FEATURE:
NAME/KEY: terminator
LOCATION: 1..836
US-08-911-434A-5

Query Match          35.2%; Score 22.2; DB 2; Length 836;
Best Local Similarity 69.8%; Pred. No. 33;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 11 AAGCACTTGTGTAGTACTTGAAGCAACCAACCAACTAG 53
Db 724 AAGCACTTATATGAACTTAAAGCCCAAAAAGAAAGTTAG 682

RESULT 6
US-08-920-812-4
Sequence 4, Application US/08920812
Patent No. 5763188
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsubisa, Akio
APPLICANT: Uehara, Hirotugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,812
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-920-812-4

Query Match 34.9%; Score 22; DB 1; Length 2885;
Best Local Similarity 63.0%; Pred. No. 50;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 1 TTCTTATTAGACCTGTGTAGTACCTTAGAAACCAACACCACTAGA 54
1639 TTGCTGATGACGATTTTATTATTAATGATGATGCCAATATACCACTATA 1632

RESULT 7
US-08-920-827-4
Sequence 4, Application US/08920827
Patent No. 570375
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,827

FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-920-827-4

Query Match 34.9%; Score 22; DB 1; Length 2885;
Best Local Similarity 63.0%; Pred. No. 50;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Db 1 TTCTTATTAGACCTGTGTAGTACCTTAGAAACCAACACCACTAGA 54
1639 TTGCTGATGACGATTTTATTATTAATGATGATGCCAATATACCACTATA 1632

RESULT 8
US-08-921-177-4
Sequence 4, Application US/08921177
Patent No. 5798211
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/921,177
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-921-177-4

Query Match 34.9%; Score 22; DB 1; Length 2885;
Best Local Similarity 63.0%; Pred. No. 50;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 1 TTCCTATTAGACCTGTGTAGTACCTTGAAACCAACACAACTTGA 54
DB 1639 TTGCTGAATGAGCATTATTATTTTAATATGATGACCAATATATACAGCTATA 1692

RESULT 9

US-08-362-577C-4
Sequence 4, Application US/08362577C
Patent No. 5807673
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsubisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,577C
FILING DATE: 27-MAR-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-362-577C-4

Query Match 34.9%; Score 22; DB 1; Length 2885;
Best Local Similarity 63.0%; Pred. No. 50;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
OY 1 TTCCTATTAGACCTGTGTAGTACCTTGAAACCAACACAACTTGA 54
DB 1639 TTGCTGAATGAGCATTATTATTTTAATATGATGACCAATATATACAGCTATA 1692

DB 1639 TTGCTGAATGAGCATTATTATTTTAATATGATGACCAATATATACAGCTATA 1692

RESULT 10

US-08-920-828-4
Sequence 4, Application US/08920828
Patent No. 5853998
GENERAL INFORMATION:
APPLICANT: Ohno, Tsuneya
APPLICANT: Matsubisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,828
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2885 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
STRAIN: Clinical Isolate SA-77
US-08-920-828-4

Query Match 34.9%; Score 22; DB 2; Length 2885;
Best Local Similarity 63.0%; Pred. No. 50;
Matches 34; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

OY 1 TTCCTATTAGACCTGTGTAGTACCTTGAAACCAACACAACTTGA 54
DB 1639 TTGCTGAATGAGCATTATTATTTTAATATGATGACCAATATATACAGCTATA 1692

RESULT 11

US-08-755-587-4
Sequence 4, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the identification and sequencing of the BRCA2 cancer

TITLE OF INVENTION: susceptibility gene and uses thereof.
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-755-587-4

Query Match 34.6%; Score 21.8; DB 3; Length 96;
Best Local Similarity 61.4%; Pred. No. 30;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 4 CTTATTAAAGCACTGTGTAGCTTGAAGAAACCAACACACACCTAGAGATCC 60
DB 23 CTTATTAAAGCACTGTGTAGCAAGATTAATGAAATCAAGAAATCTTAAAGCTTC 79

RESULT 12
US-08-755-587-28
Sequence 28, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587

FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1056 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 483..576
FEATURE:
NAME/KEY: exon
LOCATION: 481..576
US-08-755-587-28

Query Match 34.6%; Score 21.8; DB 3; Length 1056;
Best Local Similarity 61.4%; Pred. No. 48;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 4 CTTATTAAAGCACTGTGTAGCTTGAAGAAACCAACACACCTAGAGATCC 60
DB 503 CTTATTAAAGCACTGTGTAGCAAGATTAATGAAATCAAGAAATCTTAAAGCTTC 559

RESULT 13
US-08-755-587-1
Sequence 1, Application US/08755587
Patent No. 6045997
GENERAL INFORMATION:
APPLICANT: Futreal, Phillip A
APPLICANT: Wooster, Richard F
APPLICANT: Ashworth, Alan
APPLICANT: Stratton, Michael R
TITLE OF INVENTION: Materials and methods relating to the
TITLE OF INVENTION: identification and sequencing of the BRCA2 cancer
NUMBER OF SEQUENCES: 222
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson
STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
CITY: Raleigh
STATE: NC
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,587
FILING DATE: 25-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9523959.6
FILING DATE: 23-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9525555.0
FILING DATE: 14-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9617961.9

FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kenneth D Sibley
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5405-135
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1917 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-755-587-1

Query Match 34.6%; Score 21.8; DB 3; Length 1917;
Best Local Similarity 61.4%; Pred. No. 54;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 4 CTATTTAAGCACTGTGTAGTACTTGAAGAACCAACACACCACTAGAGATCC 60
DB 657 CTATTTAATGATTTTGACAGATATATGAAATCAAGAAATCTTAAGGCTTC 713

RESULT 14
US-08-840-466A-20
Sequence 20, Application US/08840466A
Patent No. 6261561

GENERAL INFORMATION:

APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Alison D.
Wachtel, Marian R.

TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.

NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:

ADDRESSEE: Flinnegan, Henderson, Parabow, Garrett &
Dunnet, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.

COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/840,466A
FILING DATE: 18-Apr-1997

CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:

NAME: Boone, Laural S.
REGISTRATION NUMBER: 43,505

REFERENCE/DOCKET NUMBER: 04995-0029-00000
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 3131 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-08-840-466A-20

Query Match 34.6%; Score 21.8; DB 4; Length 3131;
Best Local Similarity 61.4%; Pred. No. 60;

Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 1 TTCTTATTAAAGCACTGTGTAGTACTTGAAGAACCAACACCACTAGAGGA 57
DB 47 TTGGTATTACATATATGAGGAAATACATTAGAAACGACATGTGAAATAGAGGA 103

RESULT 15
US-09-696-188B-20
Sequence 20, Application US/09696188B
Patent No. 6406865

GENERAL INFORMATION:

APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Alison D.
Wachtel, Marian R.

TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Alone Or As A Fusion Protein With One Or More Other
Antigens.

NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:

ADDRESSEE: Flinnegan, Henderson, Parabow, Garrett &
Dunnet, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.

COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/696,188B
FILING DATE: 26-Oct-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/840,466
FILING DATE: 1997-04-18

ATTORNEY/AGENT INFORMATION:

NAME: Boone, Laural S.
REGISTRATION NUMBER: 43,505

REFERENCE/DOCKET NUMBER: 04995-0029-00000
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:

LENGTH: 3131 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-696-188B-20

Query Match 34.6%; Score 21.8; DB 4; Length 3131;
Best Local Similarity 61.4%; Pred. No. 60;
Matches 35; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 1 TTCTTATTAAAGCACTGTGTAGTACTTGAAGAACCAACACCACTAGAGGA 57
DB 47 TTGGTATTACATATATGAGGAAATACATTAGAAACGACATGTGAAATAGAGGA 103

Search completed: May 11, 2003, 03:07:40
Job time : 30.2632 sec

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: May 11, 2003, 03:08:16 ; Search time 62.3684 Seconds
(without alignments)
1255.289 Million cell updates/sec

Title: US-09-963-803-12
Perfect score: 63
Sequence: 1 ttccctatttaagcactgtg.....aacacactagagatccccc 63

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database:

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	100.0	63	9 US-09-963-803-12	Sequence 12, Appl
2	60	95.2	317	9 US-09-963-803-3	Sequence 3, Appl
3	60	95.2	371	9 US-09-963-803-5	Sequence 5, Appl
4	60	95.2	393	9 US-09-963-803-19	Sequence 19, Appl
5	60	95.2	462	9 US-09-963-803-20	Sequence 20, Appl
6	60	95.2	600	9 US-09-963-803-22	Sequence 22, Appl
7	49	77.8	243	9 US-09-963-803-1	Sequence 1, Appl
8	49	77.8	392	9 US-09-963-803-21	Sequence 21, Appl
9	46	73.0	301	9 US-09-963-803-7	Sequence 7, Appl
10	46	73.0	348	9 US-09-963-803-4	Sequence 4, Appl
11	46	73.0	398	9 US-09-963-803-6	Sequence 6, Appl
12	46	73.0	472	9 US-09-963-803-25	Sequence 25, Appl
13	46	73.0	541	9 US-09-963-803-24	Sequence 24, Appl
14	46	73.0	604	9 US-09-963-803-23	Sequence 23, Appl
15	25.2	40.0	373	9 US-09-796-692-6196	Sequence 6196, Ap
16	25.2	40.0	373	9 US-10-040-862-6196	Sequence 6196, Ap
17	24.8	39.4	815	10 US-09-925-300-302	Sequence 302, App
18	24.4	38.7	2915	9 US-10-028-072-209	Sequence 209, App
19	24.4	38.7	2915	9 US-10-121-049-209	Sequence 209, App

20	24.4	38.7	2915	9 US-10-123-904-209	Sequence 209, App
21	24.4	38.7	2915	9 US-10-140-470-209	Sequence 209, App
22	24.4	38.7	2915	9 US-10-175-746-209	Sequence 209, App
23	24.4	38.7	2915	9 US-10-176-918-209	Sequence 209, App
24	24.4	38.7	2915	9 US-10-176-921-209	Sequence 209, App
25	24.4	38.7	2915	9 US-10-137-865-209	Sequence 209, App
26	24.4	38.7	2915	9 US-10-140-474-209	Sequence 209, App
27	24.4	38.7	2915	9 US-10-142-431-209	Sequence 209, App
28	24.4	38.7	2915	9 US-10-143-114-209	Sequence 209, App
29	24.4	38.7	2915	9 US-10-140-002-209	Sequence 209, App
30	24.4	38.7	2915	9 US-10-142-419-209	Sequence 209, App
31	24.4	38.7	2915	9 US-10-123-262-209	Sequence 209, App
32	24.4	38.7	2915	9 US-10-142-423-209	Sequence 209, App
33	24.4	38.7	2915	9 US-10-123-050-209	Sequence 209, App
34	24.4	38.7	2915	9 US-10-141-755-209	Sequence 209, App
35	24.4	38.7	2915	9 US-10-143-032-209	Sequence 209, App
36	24.4	38.7	2915	9 US-10-123-108-209	Sequence 209, App
37	24.4	38.7	2915	9 US-10-123-436-209	Sequence 209, App
38	24.4	38.7	2915	9 US-10-123-261-209	Sequence 209, App
39	24.4	38.7	2915	9 US-10-140-921-209	Sequence 209, App
40	24.4	38.7	2915	9 US-10-140-928-209	Sequence 209, App
41	24.4	38.7	2915	9 US-10-121-045-209	Sequence 209, App
42	24.4	38.7	2915	9 US-10-123-292-209	Sequence 209, App
43	24.4	38.7	2915	9 US-10-123-903-209	Sequence 209, App
44	24.4	38.7	2915	9 US-10-124-819-209	Sequence 209, App
45	24.4	38.7	2915	9 US-10-124-822-209	Sequence 209, App

ALIGNMENTS

RESULT 1
US-09-963-803-12
; Sequence 12, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OR INVENTION: vitus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963, 803
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 63
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Directional desoxynucleotide building block
US-09-963-803-12

Query Match 100.0%; Score 63; DB 9; Length 63;
Best Local Similarity 100.0%; Pred. No. 8.4e-14;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTATTAAAGACTGTGTAGTACTTGAAGAACACACACACACCTAGAGATCC 60
DB 1 TTCTTATTAAAGACTGTGTAGTACTTGAAGAACACACACACACCTAGAGATCC 60
QY 61 CCG 63
DB 61 CCG 63

RESULT 2
US-09-963-803-3
; Sequence 3, Application US/09963803

```

Publication No. US20030028922A1
GENERAL INFORMATION
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow m
TITLE OF INVENTION: virus and cassava vein mosaic virus
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 317
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Promoter MP-1116
FEATURE:
NAME/KEY: Promoter
LOCATION: (1) ..(317)
OTHER INFORMATION:
US-09-963-803-3

```

Query Match 95.2% Score 60; DB 9; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.7e-12;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Dy 1 TTCTTATTAAAGCACTGTGTAAGTAGCTTTAGAAAACCAACAACCTTAGAGAATCC 60
| | | | |
Db 258 TTCCTATTAAAGCAGTCGTGAAGTAGCTTAGAAAAACCAACAACCTTAGAGAATCC 317

RESULT 3
US-09-963-803-5

```

1  APPLICANT: Meristem Therapeutics
2  TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
3  TITLE OF INVENTION: vitrus and casarava vein mosaic virus
4  FILE REFERENCE: 184332042
5  CURRENT APPLICATION NUMBER: US/09/963, 803
6  CURRENT FILING DATE: 2001-09-26
7  PRIOR APPLICATION NUMBER: FR 99/03925
8  PRIOR FILING DATE: 1999-03-29
9  PRIOR APPLICATION NUMBER: PCT IB00/00370
10 PRIOR FILING DATE: 2000-10-05
11 NUMBER OF SEQ ID NOS: 39
12 SOFTWARE: PatentIn version 3.1
13 SEQ ID NO 9
14 LENGTH: 371
15 TYPE: DNA
16 ORGANISM: Artificial Sequence
17 FEATURE:
18 OTHER INFORMATION: promoter Mp1146
19 FEATURE:
20 NAME/KEY: promoter
21 LOCATION: (1)..(371)
22 OTHER INFORMATION:
23 US-09-963-803-5

```

Query Match	95.2%	Score 60;	DB 9;	Length 371;
Similarity	100.0%	Pred. NO.	1.8e-12;	
Best Local				
Matches	60;	Conservative	0;	Mismatches 0;
				Gaps 0;

[illegible]

RESULT 4
US-09-963-803-19

```

1  GENERAL INFORMATION:
2  APPLICANT: Meristem Therapeutics
3  TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
4  TITLE OF INVENTION: vitruis and cassava vein mosaic virus
5  FILE REFERENCE: 184332042
6  CURRENT APPLICATION NUMBER: US/09/963.803
7  CURRENT FILING DATE: 2001-09-26
8  PRIOR APPLICATION NUMBER: FR 99/03925
9  PRIOR FILING DATE: 1999-03-29
10 PRIOR APPLICATION NUMBER: PCT IB00/00370
11 PRIOR FILING DATE: 2000-10-05
12 NUMBER OF SEQ ID NOS: 39
13 SOFTWARE: PatentIn version 3.1
14 SEQ ID NO 19
15 LENGTH: 393
16 TYPE: DNA
17 ORGANISM: Artificial Sequence
18 FEATURE:
19 OTHER INFORMATION: promoter MP1162
20 FEATURE:
21 NAME/KEY: promoter
22 LOCATION: (1)..(393)
23 OTHER INFORMATION:
24 US-09-963-803-19

```

Query Match	95.2%	Score 60	DB 9	Length 393
Best Local Similarity	100.0%	Fred. No.	1.8e-12	
Matches 60	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY 1 TTCCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACCAACCTTAGAGATCC 60

Db 334 TTCCCTATTAAAGCACTGTGTAGTACCTTAGAAACCAACCAACCTTAGAGATCC 393

RESULT 5
US-09-963-803-20

```

? APPLICANT: Meristem Therapeutics
? TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
? TITLE OF INVENTION: vitruis and cassava vein mosaic virus
? FILE REFERENCE: 184332042
? CURRENT APPLICATION NUMBER: US/09/963,803
? CURRENT FILING DATE: 2001-09-26
? PRIOR APPLICATION NUMBER: FR 99/03925
? PRIOR FILING DATE: 1999-03-29
? PRIOR APPLICATION NUMBER: PCT IB00/00370
? PRIOR FILING DATE: 2000-10-05
? NUMBER OF SEQ ID NOS: 39
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 20
? LENGTH: 462
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: promoter MP1163
? FEATURE:
? NAME/KEY: promoter
? LOCATION: (1)..(462)
? OTHER INFORMATION:
? US-09-963-803-20

```

Query Match	95.2%	Score 60;	DB 9;	Length 462;
Best Local Similarity	100.0%;	Pred. No. 1,9e-12;		
Matches 60;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 TTCCCTATTTAAGCACCTTGTGTAGTAGCTTGAAAAACCAACACAACCTAGAGGATCC 60

Db 403 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACACACTTAGAGATCC 462

RESULT 6

US-09-963-803-22
; Sequence 22, Application US/09963803
; Publication No. US2003028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1165
; NAME/KEY: promoter
; LOCATION: (1)..(600)
; OTHER INFORMATION:
US-09-963-803-22

Query Match 95.2%; Score 60; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 2,1e-12;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACACACTTAGAGATCC 60
Db 541 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACACACTTAGAGATCC 600

RESULT 7

US-09-963-803-1
; Sequence 1, Application US/09963803
; Publication No. US2003028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 243
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: 243 bp fragment from the intergenic region of commelina yellow
; OTHER INFORMATION: title virus
; NAME/KEY: promoter
; LOCATION: (1)..(243)
; OTHER INFORMATION:
US-09-963-803-1

Query Match 77.8%; Score 49; DB 9; Length 243;
Best Local Similarity 100.0%; Pred. No. 1,7e-08;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACACAC 49
Db 195 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACACAC 243

RESULT 8

US-09-963-803-21
; Sequence 21, Application US/09963803
; Publication No. US2003028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1164
; NAME/KEY: promoter
; LOCATION: (1)..(392)
; OTHER INFORMATION:
US-09-963-803-21

Query Match 77.8%; Score 49; DB 9; Length 392;
Best Local Similarity 100.0%; Pred. No. 1,9e-08;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACACAC 49
Db 334 TTCCTTATTAGACCTTGTTAGTAGCTTAGAAAAACCAACACACAC 382

RESULT 9

US-09-963-803-7
; Sequence 7, Application US/09963803
; Publication No. US2003028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 301
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1154
; NAME/KEY: promoter
; LOCATION: (1)..(301)
; OTHER INFORMATION:
US-09-963-803-7

Query Match 73.0%; Score 46; DB 9; Length 301;
Best Local Similarity 100.0%; Pred. No. 2.2e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 60
|||||
DB 256 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 301

RESULT 10
US-09-963-803-4
; Sequence 4, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 348
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1117
; NAME/KEY: promoter
; LOCATION: (1)..(348)
; OTHER INFORMATION:
US-09-963-803-4

Query Match 73.0%; Score 46; DB 9; Length 348;
Best Local Similarity 100.0%; Pred. No. 2.3e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 60
|||||
DB 303 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 348

RESULT 11
US-09-963-803-6
; Sequence 6, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1147
; NAME/KEY: promoter

; LOCATION: (1)..(398)
; OTHER INFORMATION:
US-09-963-803-6

Query Match 73.0%; Score 46; DB 9; Length 398;
Best Local Similarity 100.0%; Pred. No. 2.4e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 60
|||||
DB 353 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 398

RESULT 12
US-09-963-803-25
; Sequence 25, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MP1169
; NAME/KEY: promoter
; LOCATION: (1)..(472)
; OTHER INFORMATION:
US-09-963-803-25

Query Match 73.0%; Score 46; DB 9; Length 472;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 60
|||||
DB 427 ACTGTGTAGTAGCTTAGAAAAACCAACCAACCACTAGAGATCC 472

RESULT 13
US-09-963-803-24
; Sequence 24, Application US/09963803
; Publication No. US20030028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
; TITLE OF INVENTION: virus and cassava vein mosaic virus
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: FR 99/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; PRIOR FILING DATE: 2000-10-05
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 541
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

```

; OTHER INFORMATION: promoter MPR1168
; FEATURE:
; NAME/KEY: Promoter
; LOCATION: (1)..(541)
; OTHER INFORMATION:
US-09-963-803-24

```

```

Query Match          73.0%; Score 46; DB 9; Length 541;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 15 ACTGTGTAGTACTTGAACCAACCAACCACTAGAGATCC 60
DB 496 ACTGTGTAGTACTTGAACCAACCAACCACTAGAGATCC 541

```

```

RESULT 14
US-09-963-803-23
; Sequence 23, Application US/09963803
; Publication No. US2003028922A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics
; TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow m
; FILE REFERENCE: 184332042
; CURRENT APPLICATION NUMBER: US/09/963,803
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: PR 98/03925
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: PCT IB00/00370
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Version 3.1
; SEQ ID NO 23
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: promoter MPR1167
; NAME/KEY: promoter
; LOCATION: (1)..(604)
; OTHER INFORMATION:
US-09-963-803-23

```

```

Query Match          73.0%; Score 46; DB 9; Length 604;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 15 ACTGTGTAGTACTTGAACCAACCAACCACTAGAGATCC 60
DB 559 ACTGTGTAGTACTTGAACCAACCAACCACTAGAGATCC 604

```

```

RESULT 15
US-09-796-692-6196
; Sequence 6196, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algaier, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27

```

```

; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6196
; LENGTH: 373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (316)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-6196

```

```

Query Match          40.0%; Score 25.2; DB 9; Length 373;
Best Local Similarity 65.5%; Pred. No. 9.5;
Matches 36; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```

```

QY 5 TTATTAGACCTTGTGTAGTACTTGAACCAACCAACCACTAGAGATC 59
DB 289 TTCTTATTCATCTCTGTGAAGCTTANAAACCAACCAAGCTGTGTGCTTC 343

```

Search completed: May 11, 2003, 06:11:42
Job time : 63.3684 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 1012.74 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-12

Perfect score: 63
Sequence: 1 ttcctattacactgtf.....aacacactagagatcccg 63

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	47.6	901	12	BG294121 602390986
2	29.4	46.7	679	9	AL777174 AL777174
3	29	46.0	526	17	AG025881 AG025881
4	27.4	43.5	472	14	BQ968546 BQ968546
5	27.2	43.2	546	17	AZ652445 AZ652445
6	27	42.9	664	13	BM575510 BM575510

C	7	27	42.9	724	13	BM617342
8	26.6	42.2	514	9	AA053822	AA053822 z5f3f01.f
9	26.6	42.2	549	14	BM094741	BM094741 z5f3f01.f
10	26.6	42.2	843	14	BQ963329	BQ963329 AGENCOURT
11	26.4	41.9	343	12	BG461281	BG461281 RST44060
12	26.4	41.9	501	17	AZ407729	AZ407729 IM0178002
13	26.2	41.6	389	9	AA757510	AA757510 z5f3f01.f
14	26.2	41.6	750	13	BI917143	BI917143 603181563
15	26.2	41.6	762	14	BQ745595	BQ745595 UI-M-EMO-
16	25.8	41.0	762	14	BH714488	BH714488 BOMDC72RF
17	25.6	40.6	458	17	AG685997	AG685997 HS_5550_A
18	25.6	40.6	509	12	BF156127	BF156127 F17H10.Y
19	25.6	40.6	540	17	AQ804388	AQ804388 HS_3248_A
20	25.6	40.6	549	13	BM037729	BM037729 S096H06.S
21	25.6	40.6	559	13	BM155751	BM155751 fw26H07.Y
22	25.6	40.6	662	14	BQ616770	BQ616770 faa94h07
23	25.6	40.6	677	13	BI903774	BI903774 603167748
24	25.6	40.6	708	17	AQ898421	AQ898421 HS_3135_B
25	25.6	40.6	716	10	AV705148	AV705148 AV705148
26	25.4	40.3	492	17	AZ312818	AZ312818 IM0028G24
27	25.4	40.3	638	17	AZ248833	AZ248833 RPT-23-9
28	25.4	40.3	858	12	BG617584	BG617584 602615075
29	25.4	40.3	911	12	BF794977	BF794977 602256877
30	25.2	40.0	380	14	BQ185082	BQ185082 UI-E-EJL-
31	25.2	40.0	771	17	CNS05RVV	AL351076 Tetraodon
32	25	39.7	459	10	AM085143	AM085143 x04906.x
33	25	39.7	534	17	BH823752	BH823752 BACP2-D0
34	25	39.7	817	12	BE884936	BE884936 601506870
35	25	39.7	931	17	AZ539154	AZ539154 ENT0547F
36	25	39.7	931	17	AZ539154	BQ39668 AGENCOURT
37	25	39.7	1220	14	BG467453	BG467453 Na_L3_04A
38	24.8	39.4	264	12	R34801	R34801 yb87602-r1
39	24.8	39.4	286	14	BG895557	BG895557 359062.MA
40	24.8	39.4	302	12	BQ488731	BQ488731 68-B8456-
41	24.8	39.4	367	14	BM784926	BM784926 K-EST0063
42	24.8	39.4	378	12	BF079516	BF079516 230139.MA
43	24.8	39.4	440	10	AW781100	AW781100 s189a10.Y
44	24.8	39.4	457	17	AQ225617	AQ225617 HS_2003_B

ALIGNMENTS

RESULT 1
LOCUS BG294121/c 901 bp mRNA linear EST 21-FEB-2001
DEFINITION 602390986F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4502925 5',
ACCESSION BG294121
VERSION BG294121.1 GI:13054439
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
Tissue Procurement: The Cepo Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LLNL0372 row: k column: 22
High quality sequence stop: 718.
Location/Qualifiers 1..901

FEATURES
source

/organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4502925"
 /clone_1lb="NIH_MGC_94"
 /libuse_type="retina"
 /lib_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dt primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC library."

BASE COUNT 234 a 195 c 204 g 268 t

Query Match 47.6%; Score 30; DB 12; Length 901;
 Best Local Similarity 67.7%; Pred. No. 9.3;

Matches 42; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 1 TTCCTATTAGACCTGTGTAGTACCTTAGAAAACCAACACACACCTAGAGATCC 60
 Db 266 TTCACTATCCAAACACGAGTAGGTGTGAAACCAAAACCCCACTTAAGAACCA 207
 Qy 61 CC 62
 Db 206 CC 205

RESULT 2

AL777174/c AL777174 679 bp mRNA linear EST 25-JUN-2002
 LOCUS AL777174 XGC-gastrula silurana tropicalis cDNA clone TGa8081f17 5',
 DEFINITION mRNA sequence.

ACCESSION AL777174 GI:21562878
 VERSION AL777174.1

KEYWORDS EST.
 SOURCE western clawed frog.

ORGANISM Silurana tropicalis
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopodidae; Silurana.

REFERENCE 1 (bases 1 to 679)
 AUTHORS Taylor,R., Ashurst,J.L., Cronling,M.D.R., Zorn,A.M. and Rogers,J.
 TITLE Sanger Xenopus tropicalis EST project 2002
 JOURNAL Unpublished (2001)
 COMMENT Contact: Taylor R

Sanger Centre
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS SEQUENCE ID: TGa8081f17.plksp6
 Sequencing primer: PKSP6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.

FEATURES

Location/Qualifiers

1..679
 /organism="Silurana tropicalis"
 /db_xref="taxon:8364"
 /clone="TGa8081f17"
 /clone_1lb="XGC-gastrula"
 /dev_stage="gastrula (stages 10.5-13 mixed)"
 /lab_host="Escherichia coli XL1-blue"
 /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
 was oligo dt primed from 5' end of poly A+ RNA from stages
 10-13 gastrulae. EcoRI/NotI cut cDNA was then ligated
 into pCS107 with EcoRI at the 5' end and NotI at the 3'
 end."

BASE COUNT 213 a 134 c 116 g 216 t

Query Match 46.7%; Score 29.4; DB 9; Length 679;
 Best Local Similarity 70.9%; Pred. No. 15;
 Matches 39; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 4 CTTATTAGACCTGTGTAGTACCTTAGAAAACCAACACACCACTAGAGAT 58
 Db 648 CTATTCAAACCTGTGTGTAAATATTAACCAACCAATTACACTATAGAT 594

RESULT 3

AG025881 526 bp DNA linear GSS 21-DEC-1999
 LOCUS Homo sapiens DNA, chromosome 8q23, reverse end of BAC clone:
 DEFINITION KB1562D12, genomic survey sequence.

ACCESSION AG025881 GI:6624572
 VERSION AG025881.1

KEYWORDS GSS.
 SOURCE Homo sapiens DNA, clone_1lb:KEIO BAC library clone:KB1562D12.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 526)
 AUTHORS Shimizu,N. and Asakawa,S.
 TITLE The BAC end sequence of Homo sapiens BAC clone KB1562D12
 JOURNAL Published only in Database (1999)
 REFERENCE 2 (bases 1 to 526)
 AUTHORS Shimizu,N. and Asakawa,S.
 TITLE Direct Submission
 JOURNAL Submitted (15-DEC-1999) Nobuyoshi Shimizu, Keio University, School
 of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo
 160-8582, Japan (E-mail:nshimizu@med.keio.ac.jp.
 Tel:81-3-3351-2370, Fax:81-3-3351-2370)

FEATURES

Location/Qualifiers

1..526
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="8"
 /map="8q23"
 /clone="KB1562D12"
 /clone_1lb="KEIO BAC library"
 /note="This sequence is reverse end of BAC clone
 KB1562D12."

BASE COUNT 112 a 146 c 90 g 178 t

Query Match 46.0%; Score 29; DB 17; Length 526;
 Best Local Similarity 71.7%; Pred. No. 20;
 Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 8 TTTAAGACCTGTGTAGTACCTTAGAAAACCAACACCAACCTAGAGATCC 60
 Db 343 TTAAACACTTCAGTGTGCTGCTGAGATTCTACACACACCTGAAGATCC 395

RESULT 4

BO968546/c BO968546 472 bp mRNA linear EST 21-AUG-2002
 LOCUS OHB34E11.yg.ab1 OH ABCDI sunflower RH801 Helianthus annuus cDNA
 DEFINITION clone OHB34E11, mRNA sequence.

ACCESSION BO968546 GI:22386067
 VERSION BO968546.1

KEYWORDS EST.
 SOURCE common sunflower.

ORGANISM Helianthus annuus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
 Heliantheae; Helianthus.

REFERENCE 1 (bases 1 to 472)
 AUTHORS Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison,
 P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
 Church,S., Jackson,L. and Bradford,K.
 TITLE Lettuce and Sunflower ESTs from the Composite Genome Project
 JOURNAL http://compgenomics.ucdavis.edu/
 COMMENT Unpublished (2002)
 CONTACT: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W. Michelmore Lab
University of California at Davis (UCD)
Assumeson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QH_Ca_Contig1526, see http://cgdb.ucdavis.edu/
for details.
Plate: OHR34
Location/Qualifiers
1..472
/organism="Helianthus annuus"
/cultivar="RHA801"
/db_xref="taxon:4232"
/clone="OHR34E11"
/clone_1lb="OH ABCDI sunflower RHA801"
/lab_host="E.coli"
/note="Vector: pERCDSAF1AB; The library was constructed from 11 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/TAG_SEQ=Not found"

BASE COUNT 124 a 81 c 108 g 155 t 4 others
ORIGIN

Query Match 43.5%; Score 27.4; DB 14; Length 472;
Best Local Similarity 68.5%; Pred. No. 67;
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 TTCTTATTAGACCTGTGTAGCTTGAAGAAACCAACACACACTTGA 54
77 TTACATCATTAACACCTTTGAGATGATTAACCAATATGAGAGAGA 24

RESULT 5
LOCUS A2652445/c 546 bp DNA linear GSS 14-DEC-2000
DEFINITION IM0525J11R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0525J11 R, DNA sequence.
ACCESSION A2652445
VERSION A2652445.1 GI:11788967
KEYWORDS GSS.
SOURCE mouse mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
1 (bases 1 to 546)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0525 row: J column: 11
Seq primer: CACACAGAAACACACTTATGACC
Class: plasmid ends
High quality sequence stop: 546.
Location/Qualifiers

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES

source
1..546
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0525J11"
/clone_1lb="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42ny; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The ligated DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 133 a 105 c 132 g 176 t
ORIGIN

Query Match 43.2%; Score 27.2; DB 17; Length 546;
Best Local Similarity 72.9%; Pred. No. 78;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 TTCTTATTAGACCTGTGTAGCTTGAAGAAACCAACACACAGA 48
Db 217 TACCATTTTGAGAAATTGTTATCATCAAGAAACCAACCAACAA 170

RESULT 6
LOCUS BM575510/c 664 bp mRNA linear EST 22-FEB-2002
DEFINITION 17000687161198 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
19600449696818 5', mRNA sequence.
ACCESSION BM575510
VERSION BM575510.1 GI:18863977
KEYWORDS EST.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
1 (bases 1 to 664)
Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
Celera Anopheles gambiae EST project
Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404531151
Fax: 2404534580
Email: HoltR@celera.com
Plate: ND01004AAV row: D column: 20
Seq primer: M13 Reverse.
Location/Qualifiers

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES

```

/dev stage="Adult"
/lab host="DH10b"
/notes=vector: DSPORT; Site_1: SalI; Site_2: NotI; Whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. CDNA inserts >500 bp
cloned directionally into DSPORT 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mr4.org)

```

Query Match	42.9%	Score 27	DB 13	Length 664
Best Local Similarity	70.6%	Pred. No. 91		
Matches	36	Conservative	0	Mismatches 15; Indels 0; Gaps 0;
7	ATTTAGACCTGTGTAGTAAAGAAAACCAACACAACTACAGGA	57		
DB	526	ATTTAGCTCTGTGCAGTAAACAAATATACAAACAACTACAGCA	476	

RESULT 7	BM617342/c	LOCUS	DEFINITION	ACCSSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	BM617342	17000687515386 A.Gam.ad.cDNA.bloodi	724 bp mRNA linear EST 25-FEB-2002					
		19600449791520 5', mRNA sequence.						
	BM617342							
	BM617342.1	GI:18915564						
	EST.							
	African malaria mosquito.							
	Anopheles gambiae							

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (Pages 1 to 724)	Holt, R.A., Lin, J.-J., Murphy, S. D., Evans, C.A., Kraft, C.L., Charlalab, R., Colling, F.H., Venter, J.C. and Hoffman, S.L.	Cefera Anopheles gambiae EST project	Unpublished (2002)	Contact: Holt R.A.

```
FEATURES      Email: HOITRA@celera.com  
              Plate: N01I0049UR row: K column: 10  
              Seq primer: M13 Reverse.  
              Location/Qualifiers  
source        1..724
```

```

/organism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/ld_xref="taxon:7165"
/clone="1960049731920"
/clone_1ib="A. Gam. ad. CDNA, blood1"
/clone_stage="Adult"
/lab_host="DHI10b"
/notes="Vector: pSport1; Site 1: SalI; Site 2: NotI; whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. CDNA inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mrt4.org)"

```

Query Match	42.9%	Score 27	DB 13	Length 724
Best Local Similarity	70.6%	Pred. No. 92		
Matches	36	Conservative	0	Mismatches 15, Indels 0, Gaps 0

Oy	7	ATTTAAAGCACTTGTGTAGTACCTTACAAAACCAACAAACACCTTAGAGA	57
Db	92	AATTAGGCTCTGTGTGAGTAAACAAATATAACAAACAAACAACTTAGACA	42

RESULT 8	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE
AA053822	514 bp	linear	EST 02-FEB-1997		
AA053822	z553f01.r1	Soares retina N2bHR	Homologous	CDNA clone	
AA053822	IMAGE:380665	5', mRNA sequence.			
AA053822.1	GI:1544757				
EST.					
human.					

REFERENCE 1 (Pages 510 to 514)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Donald, M. F., Chiappelli, B.,
 Chisese, S., Dietrich, N., Dubucque, T., Faveello, A., Gish, W., Hawkins,
 M., Hulman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Maridis, B., Moore,
 S., Morris, M., Parsons, J., Prange, C., Rickin, L., Rohlfing, T.,
 Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Trevisakis, E.,
 Underwood, K., Wohlmann, P., Waterson, P., Wilson, R. and Merra, M.
 TITLE Generation and analysis of 200,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE
 COMMENT
 Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Putative full length read
The vector to vector length is 1104
Insert length: 2734 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence, stop: 452.

```

FEATURES
SOURCE
LOCATION/Qualifiers
1..514
/organism="Homo sapiens"
/db_xref="GDB:1288922"
/db_xref="taxon:3606"
/clone="IMAGE:380665"
/clone_1p="Soares retina N2b4HR"
/sex="male"

```

BASE COUNT
ORIGIN

136 a 121 c 117 g 138 t 2 others

Query Match	42.2%	Score 26.6	DB 9	Length 514
Best Local Similarity	66.7%	Pred. No. 1.2e+07		
Matches 38	Conservative 0	Mismatches 19	Indels 0	Gaps 0
Qy	7	ATTTAAGCACTTGTGTAGCTTAGAAAAACAACACAACTAGAGGATCCCCG	63	
Db	347	ATTATATGCTCTGTGTAGAGAGCTAGAAAAAAGAAAGAAAAAGAACTCTGAG	403	

RESULT 9

LOCUS	BM094741	549 bp	mRNA	linear	EST 30-NOV-2001
DEFINITION	8a320e07.v1 Gm-c1066 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1066-3301 5' similar to TR:Q9SH35 Q9SH35 F2K1.13. // mRNA Sequence.				
ACCESSION	BM094741				
VERSION	BM094741.1	GI:17023707			
KEYWORDS	EST.				
SOURCE	Soybean.				
ORGANISM	Glycine max				
REFERENCE	1 (bases 1 to 549) Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurr, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R. Public Soybean EST Project ; Unpublished (1999)				
TITLE	Contact: Shoemaker R/Public Soybean EST Project				
JOURNAL	Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu				
COMMENT	Possible reversed clone; similarity on wrong strand This clone is available through: Resgen, Invitrogen Corp. 2110 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact via email: coudre@resgen.com				
FEATURES	High quality sequence stop: 421.				
SOURCE	Location/Qualifiers				
	1..549				
	/organism="Glycine max"				
	/db_xref="taxon:3847"				
	/clone="GENOME SYSTEMS CLONE ID: Gm-c1066-3301"				
	/clone_id="Gm-c1066"				
	/tissue_type="leaf and shoot tip, salt stressed, 2 week old seedling"				
	/lab host="DH10B"				
	/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from unexpanded leaves and the shoot tips of 2 week old seedlings from the cultivar Williams. The 2 week old seedlings were salt stressed in a solution of 500mM NaCl for 3 days prior to harvesting. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."				
BASE COUNT	187 a	124 c	82 g	156 t	
ORIGIN					
Query Match	42.2%; Score 26.6; DB 13; Length 549;				
Best Local Similarity	71.4%; Pred. No. 1.2e+02;				
Matches 35; Conservative	0;	Mismatches 14;	Indels 0;	Gaps 0;	
1	TTTCCTATTATTAAGCACTTGCTAGTAGCTTGAAGAAACAACAACAAC	49			
Db	9	TTTTTTTATTAAGCAATTGCTCTGTCATTAATTAATTAAGCAACAACAAC	57		
RESULT 10					
R0963329/c					

LOCUS	BQ6963329		843 bp	mRNA	linear	EST 21-APR-2002
DEFINITION	AGENCOURT_10050135 NIH_MGC_114 Mus musculus cDNA clone IMAGE:6508493 5', mRNA sequence.					
ACCESSION	BQ6963329					
VERSION	BQ6963329.1		GI:22378807			
KEYWORDS	EST.					
SOURCE	house mouse.					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.					
AUTHORS	1 (bases 1 to 843)					
TITLE	NIH-MGC http://mgs.nci.nih.gov/.					
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)					
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaps@remail.nih.gov Tissue Procurement: Dr. David Rowe cDNA Library Preparation: Invitrogen Corp CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLT at: http://image.lnl.gov Plate: LRAM14074 row: e column: 06 High quality sequence stop: 467.					
FEATURES	Location/Qualifiers					
source	1..843					
	/organism="Mus musculus"					
	/db_xref="taxon:10090"					
	/clone="IMAGE:6508493"					
	/clone_id="NIH_MGC_134"					
	/tissue_type="undifferentiated limb"					
	/lab_host="DH10B (phage-resistant)"					
	/note="Vector: pCMV-SPORT6.1.ccdp; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Constructed by Resgen, Invitrogen Corp. Note: This is a NIH_MGC Library."					
BASE COUNT	214 a 204 c 191 g 233 t					1 others
ORIGIN						
Query Match	42.2%; Score 26.6; DB 14; Length 843;					
Best Local Similarity	66.7%; Pred. No. 1.2e+02;					
Matches	38; Conservative 0; Mismatches 19; Indels 0; Gaps 0;					
Dn	6 TATTTAAGCATTCTTGTTGTAGTAGCTTGAAGAAACAACAACAACACTGAGATGCC 62					
	142 TATGTGTCACTGTTCATTACACACAGAACCMAAACAACAACCCAGTATGCC 86					
RESULT 11						
LOCUS	BG461281		343 bp	mRNA	linear	EST 21-APR-2001
DEFINITION	R5744060 Atherys RAGE Library Homo sapiens cDNA, mRNA sequence.					
ACCESSION	BG461281					
VERSION	BG461281.1		GI:13749787			
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS	1 (bases 1 to 343)					
	Harrington,D.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R., Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J., Lerner,L., Costanzo,D., McElligott,K., Booser,S., Maye,R., Smith,E., Veloso,N., Kikta,A., Hess,D., Cotren,K., Lo,K., Offenbacher					
TITLE	'J., Danzig,J. and Ducar,M.					
JOURNAL	Creation of genome-wide protein expression libraries using random					
MEBLINE	Nat. Biotechnol. 19 (5), 440-445 (2001)					
COMMENT	21227151					
	Contact: Scott J. Cain					
	Atherys, Inc.					
	3201 Carnegie Ave, Cleveland, OH 44115, USA					

Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@etherys.com
Location/Qualifiers

FEATURES

SOURCE

1. 343
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 135 a 81 c 46 g 81 t
ORIGIN

Query Match 41.9%; Score 26.4; DB 12; Length 343;
Best Local Similarity 69.2%; Pred. No. 1.4e+02;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 7 ATTTAGCACTGTGTAGTCTTGAAGAACCAACCAACCTAGAGAT 58
Db 4 ATCAAGACTGTATATATATTTGGAACCAACCAACCAAGAGAT 55

RESULT 12 501 bp DNA linear GSS 03-OCT-2000
AAZ407729
LOCUS
DEFINITION
clone178002R Mouse 10kb plasmid UGCM1 library Mus musculus genomic
AAZ407729
ACCESSION
AAZ407729.1 GI:10531742

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
1 (bases 1 to 501)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingay, A., von Niederhausern, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddu@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0178 row: 0 column: 02
Seq primer: CACACAGAAACACCTGTGACC
Class: plasmid ends
High quality sequence stop: 501.
Location/Qualifiers

FEATURES

SOURCE

1. 501
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UGCM178002"
/clone_lib="Mouse 10kb plasmid UGCM1 library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Ti-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 116 a 112 c 125 g 148 t
ORIGIN

Query Match 41.9%; Score 26.4; DB 17; Length 501;
Best Local Similarity 69.2%; Pred. No. 1.4e+02;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 7 ATTTAGCACTGTGTAGTCTTGAAGAACCAACCAACCTAGAGAT 58
Db 422 ATTTAGCACTGTGTCTTGTGGAAGAACCAAGCAAGCTTGAAGCTT 473

RESULT 13 389 bp mRNA linear EST 23-JAN-1998
AA757510
LOCUS
DEFINITION
2937809.s1 Soares pineal gland N3HPG Homo sapiens cDNA clone
IMAGE:395512 3', mRNA sequence.
AA757510
ACCESSION
AA757510.1 GI:2805373

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 389)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S., Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Mair, M., Martin, J., Moore, B., Schellberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterson, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 353.
Location/Qualifiers

FEATURES

SOURCE

1. 389
/organism="Homo sapiens"
/db_xref="GDB:1301414"
/db_xref="taxon:9606"
/clone_lib="IMAGE:395512"
/clone_lib="Soares pineal gland N3HPG"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: pineal gland; Vector: pTR73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' GTTACCAATCTGAAGTGGAGCGCGCGCTTTTCTTTTCTTTT 3'] double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTR73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT	107 a	79 c	69 g	134 t		
ORIGIN						
Query Match	41.6%	Score 26.2;	DB 9;	Length 389;		
Best Local Similarity	67.3%;	Pred. No. 1.7e+02;				
Matches 37; Conservative	0;	Mismatches 18;	Indels 0;	Gaps 0;		
Oy	7	ATTATGACCTGTGTAGTCTTGAACCAACACCAACACTAGAGATCC	61			
Db	77	ATTATGCTCTGTGTAGAGCTGAGAAAAAAGAGAAAAAAGAAAGAACTCTC	23			
RESULT 14						
LOCUS	B1917143	750 bp	mRNA	linear		
DEFINITION	603181563p1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5245646 5',					
ACCESSION	B1917143					
VERSION	B1917143.1	GI:16181105				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
TITLE	1 (bases 1 to 750)					
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.					
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)					
	Unpublished (1999)					
	Contact: Robert Strausberg, Ph.D.					
	Email: cga@bbs-rcmail.nih.gov					
	Tissue Procurement: Life Technologies, Inc.					
	cDNA Library Preparation: Life Technologies, Inc.					
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)					
	DNA Sequencing by: Incyte Genomics, Inc.					
	Clone distribution: MGC clone distribution information can be					
	found through the I.M.A.G.E. Consortium/LNLN at:					
	http://image.lnl.gov					
	Plate: LLM11619	row: n	column: 15			
FEATURES	High quality sequence stop: 748.					
source	Location/Qualifiers					
	1..750					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="IMAGE:5245646"					
	/clone_id="NIH_MGC_121"					
	/lab_host="DH10B"					
	/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;					
	Site 2: EcoRV (destroyed); RNA source anonymous pool of 3					
	fetal brains, female age 20 weeks, female age 24 weeks,					
	and male age 26 weeks. Library is oligo-dT primed and					
	directionally cloned (EcoRV site is destroyed upon					
	cloning). Average insert size 1.7 kb, insert size range					
	0.7-3.5 kb. Library is normalized and enriched for					
	full-length clones and was constructed by C. Gruber					
	(Invitrogen). Research Genetics tracking code 017. Note:					
	this is a NIH-MGC Library."					
BASE COUNT	225 a	139 c	170 g	216 t		
ORIGIN						
Query Match	41.6%	Score 26.2;	DB 13;	Length 750;		
Best Local Similarity	67.3%;	Pred. No. 1.7e+02;				
Matches 37; Conservative	0;	Mismatches 18;	Indels 0;	Gaps 0;		
Oy	5	TTATTTAAGCACTGTGTAGTCTTGAACCAACCAACAACTAGAGATC	59			
Db	431	TTCTTATTCACCTCTGTGGAAGCTGAGAAACCAAAACACACTGTGTGCTTC	485			
RESULT 15						
LOCUS	B0745595	762 bp	mRNA	linear		
DEFINITION	UI-M-EMO-bxb-g-17-0-UI.r1 NIH_MGC_Mus musculus cDNA clone					
	IMAGE:5708200 5', mRNA sequence.					

```

ACCESSION      BO745595
VERSION        BO745595.1
KEYWORDS       GI:21892382
SOURCE         EST.
ORGANISM       house mouse.
               Mus musculus.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 762)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgabs-remail.nih.gov
               Tissue Procurement: Dr. James Lin, University of Iowa
               CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
               CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
               DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
               Clone Distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LNLW at:
               http://image.lnl.gov
               This clone was contributed by the Brain Molecular Anatomy Project
               (BMAP)

FEATURES
   source
       location/Qualifiers
           1..762
               /organism="Mus musculus"
               /strain="C57BL/6"
               /db_xref="taxon:10090"
               /clone="IMAGE:5708200"
               /clone_id="NIH_BMAP_EW0"
               /tissue_type="whole brain"
               /dev_stage="embryo 15.5 dpc"
               /lab_host="DH10B (T1 phage resistant)"
               /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
               Site_2: Not I; The library was constructed according to
               Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
               1996. Denatured mRNA was size fractionated on a 1% agarose
               gel. First strand cDNA synthesis was primed with an
               oligo-dT primer containing a Not I site. Double stranded
               cDNA was size selected according to mRNA size fraction,
               ligated with EcoR I adaptor, digested with Not I, and then
               cloned directionally into pYX-Asc vector. The library tag
               sequence located between the Not I site and the polyA tail
               is GNGCGTGGA. This library was created for the
               University of Iowa Mouse Brain Molecular Anatomy Project
               (BMAP): 'Gene Discovery in the Developing Mouse Nervous
               System', supported by National Institutes of Mental Health
               (NIMH), Hemin Chn, Ph.D., program coordinator."
BASE COUNT     171 a      148 c      198 g      244 t      1 others
ORIGIN
Query Match    41.6%; Score 26.2; DB 14; Length 762;
Best Local Similarity 67.3%; Pred. No. 1.7e+02;
Matches 37; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
OY             9 TTAAGCACTTGAGTAGCTTAGAAAACCAACACACAACTTAGAGATCCCG 63
Db             490 TCAAGCACTTAATTAATTAATTCACAAAACCAACACCAACCCATCATGCACG 436

Search completed: May 11, 2003, 04:54:18
Job time : 1017.74 secs

```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:18:55 ; Search time 153.644 Seconds
(without alignments)
1157.922 Million cell updates/sec

Title: US-09-963-803-14

Perfect score: 79
Sequence: 1 catgcgcagactatgcat.....aaggatgacatgcacact 79

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq.101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	79	21	AAA96848
2	79	100.0	79	21	AAA96472
3	75	94.9	259	21	AAA96477
4	75	94.9	259	21	AAA96465
5	70	88.6	472	21	AAA96859
6	70	88.6	541	21	AAA96858
7	69.6	88.1	393	21	AAA96853
8	69.6	88.1	462	21	AAA96854
9	69.6	88.1	600	21	AAA96856

C	10	69	87.3	392	21	AAA96855	Nucleotide sequenc
	11	68.2	86.3	604	21	AAA96857	Nucleotide sequenc
	12	68	86.1	371	21	AAA96839	Nucleotide sequenc
C	13	58	73.4	80	22	AA03420	CanV as-2/as-1 PCR
	14	58	73.4	226	22	AA03396	Mpr1133 promoter D
	15	58	73.4	229	22	AA03397	Mpr1133 promoter D
	16	58	73.4	332	22	AA03398	Mpr1133 promoter D
	17	58	73.4	472	22	AA03394	Mpr1130 promoter D
	18	55	69.6	80	22	AA03421	CanV as-2/as-1 PCR
	19	46	58.2	63	21	AAA96847	Directional desoxy
	20	46	58.2	63	21	AAA96471	Nucleotide sequenc
	21	46	58.2	280	21	AAA96463	Nucleotide sequenc
	22	46	58.2	301	21	AAA96841	Nucleotide sequenc
	23	46	58.2	398	21	AAA96840	Nucleotide sequenc
C	24	39.2	49.6	63	22	AA03422	CanV as-2/as-1 PCR
	25	39.2	49.6	219	22	AA03399	Mpr1136 promoter D
	26	39.2	49.6	282	22	AA03400	Mpr1137 promoter D
	27	39.2	49.6	315	22	AA03401	Mpr1138 promoter D
	28	39.2	49.6	381	22	AA03406	Mpr1200 promoter D
	29	39.2	49.6	455	22	AA03395	Mpr1131 promoter D
	30	39.2	49.6	505	22	AA03402	Mpr1139 promoter D
	31	39	49.4	189	22	AA01006	Amplified product
	32	39	49.4	240	21	AAA71895	Soybean RRS gene N
	33	39	49.4	309	21	AA087192	CanV35S promoter e
	34	39	49.4	309	24	AA039442	Caulliflower mosaic
	35	39	49.4	333	22	AA03764	Plasmid pMON295 Ca
	36	39	49.4	334	22	AA028369	Nucleotide sequenc
	37	39	49.4	338	20	AA099492	Caulliflower mosaic
	38	39	49.4	350	13	AA023533	CanV35S promoter (
	39	39	49.4	352	13	AA025080	CanV35S promoter,
	40	39	49.4	395	15	AA072688	Caulliflower mosaic
	41	39	49.4	396	17	AA032299	Caulliflower mosaic
	42	39	49.4	396	22	AA028368	Nucleotide sequenc
	43	39	49.4	400	20	AA099491	Sequence containin
	44	39	49.4	444	20	AA086466	35S CMV promoter s
	45	39	49.4	444	22	AA089658	Tomato spotted wil

ALIGNMENTS

RESULT 1	
AAA96848	AAA96848 standard; DNA; 79 BP.
ID	AAA96848
AC	AAA96848;
XX	
DT	19-FEB-2001 (first entry)
XX	
DE	Directional desoxynucleotide building block 57.
XX	
KW	Promoter, intergenic region; Comelina yellow mottle virus;
KW	chimeric expression promoter; plant vascular expression promoter;
KW	plant green tissue expression promoter; Cassava vein mosaic virus;
KW	transgenic plant; ss.
XX	
OS	Synthetic.
XX	
PN	WO200058485-A1.
XX	
PD	05-OCT-2000.
XX	
PF	29-MAR-2000; 2000WO-IB00370.
XX	
PR	29-MAR-1999; 99FR-0003925.
XX	
PA	(MERI-) MERISTEM THERAPEUTICS.
XX	
PI	Rance I, Gruber V, Theisen M;
XX	
DR	WPI; 2000-647238/62.
XX	
PT	Chimeric expression promoter for transgenic plant production, comprises

PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
PS
XX Disclosure; Page 26; 91pp; English.
XX
CC The present sequence represents a directional deoxynucleotide building
CC block, which was used to construct chimeric promoters of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Comelina yellow mottle virus, and the second plant promoter originates
CC from the Casarva vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 79 BP; 22 A; 15 C; 20 G; 22 T; 0 other;
Query Match 100.0%; Score 79; DB 21; Length 79;
Best Local Similarity 100.0%; Pred. No. 5.2e-20;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGCTGAGACTAGTATGATGATATCAAGATGATGATATCTCCAGTACGTA 60
DB 1 CATGCTGAGACTAGTATGATGATATCAAGATGATGATATCTCCAGTACGTA 60
QY 61 AGGATGACGATGCCACT 79
DB 61 AGGATGACGATGCCACT 79
RESULT 2
AAA96472
ID AAA96472 standard; DNA; 79 BP.
XX
AC AAA96472;
XX
DT 08-FEB-2001 (first entry)
XX
DE Nucleotide sequence of the directional building block 57.
XX
KM petE promoter; chimeric promoter; transgenic plant; MPr1108;
KM plastocyanin gene promoter; PCR primer; 88.
XX
OS Synthetic.
XX
OS WO200056906-A1.
XX
PN 28-SEP-2000.
PD
XX 20-MAR-2000; 2000WO-IB00317.
PF
XX 22-MAR-1999; 99FR-0003635.
PR
XX (MERT-) MERISTEM THERAPEUTICS.
PA
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-587667/55.
DR
XX
PT Chimeric expression promoter for producing dicotyledonous and
PT monocotyledonous transgenic plants comprises a nucleic acid sequence
PT derived from a promoter of the pea plastocyanin gene
XX
XX
PS Claim 24; Page 76; 83pp; English.
XX
CC The present sequence represents the directional building block 57,
CC which is used to construct chimeric promoters of the invention in
CC PCR reactions. The specification describes a chimeric expression
CC promoter comprising a petE promoter of the pea plastocyanin gene, or

CC comprising a G box operably or functionally linked upstream of a
CC CAAT box, TATA box and transcription initiation site. The chimeric
CC promoters are used in expression vectors for producing transgenic
CC plants, such as dicotyledonous species, e.g. potato, tobacco, cotton,
CC lettuce, tomato, melon, cucumber, pea, rape, beetroot, or sunflower,
CC and monocotyledonous species, e.g. wheat, barley, oat, rice, or corn.
XX
SQ Sequence 79 BP; 22 A; 15 C; 20 G; 22 T; 0 other;
Query Match 100.0%; Score 79; DB 21; Length 79;
Best Local Similarity 100.0%; Pred. No. 5.2e-20;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGCTGAGACTAGTATGATGATATCAAGATGATGATATCTCCAGTACGTA 60
DB 1 CATGCTGAGACTAGTATGATGATATCAAGATGATGATATCTCCAGTACGTA 60
QY 61 AGGATGACGATGCCACT 79
DB 61 AGGATGACGATGCCACT 79
RESULT 3
AAA96477
ID AAA96477 standard; DNA; 259 BP.
XX
AC AAA96477;
XX
DT 08-FEB-2001 (first entry)
XX
DE Nucleotide sequence of the promoter MPr1112.
XX
KM petE promoter; chimeric promoter; transgenic plant; MPr1112;
KM plastocyanin gene promoter; 88.
XX
OS Pisum sativum.
XX
OS WO200056906-A1.
XX
PN 28-SEP-2000.
PD
XX 20-MAR-2000; 2000WO-IB00317.
PF
XX 22-MAR-1999; 99FR-0003635.
PR
XX (MERT-) MERISTEM THERAPEUTICS.
PA
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-587667/55.
DR
XX
PT Chimeric expression promoter for producing dicotyledonous and
PT monocotyledonous transgenic plants comprises a nucleic acid sequence
PT derived from a promoter of the pea plastocyanin gene
XX
XX
PS Claim 25; Page 79; 83pp; English.
XX
CC The present sequence represents the chimeric promoter MPr1112. The
CC promoter is derived from the petE promoter from pea plastocyanin gene,
CC by fusing the petE as-1 like and nos enhancer like elements to the
CC promoter MPr1098 (comprising TATA and CAAT boxes of petE), and then
CC fusing a fragment comprising a duplication of the element as2 and as1.
CC The petE promoter directs cell-specific but not full light-regulated
CC expression in transgenic tobacco plants. The promoter is used to
CC construct chimeric promoters of the invention. The specification
CC describes a chimeric expression promoter comprising a promoter of
CC the pea plastocyanin gene, or comprising a G box operably or
CC functionally linked upstream of a CAAT box, TATA box and transcription
CC initiation site. The chimeric promoters are used in expression vectors
CC for producing transgenic plants, such as dicotyledonous species,
CC e.g. potato, tobacco, cotton, lettuce, tomato, melon, cucumber, pea,
CC rape, beetroot, or sunflower, and monocotyledonous species, e.g. wheat,
CC barley, oat, rice, or corn.

XX Sequence 259 BP; 82 A; 67 C; 38 G; 72 T; 0 other;
SQ
Query Match 94.9%; Score 75; DB 21; Length 259;
Best Local Similarity 100.0%; Pred.No. 2.4e-18;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 CTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAAAGG 64
DB 13 CTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAAAGG 72
QY 65 ATGACGATGCCACT 79
DB 73 ATGACGATGCCACT 87
RESULT 4
AAA96465
ID AAA96465 standard; DNA; 296 BP.
XX
AC AAA96465;
XX
DT 08-FEB-2001 (first entry)
XX
DE Nucleotide sequence of the promoter MP1111.
XX
KM petB promoter; chimeric promoter; transgenic plant; MP1111;
KW plastocyanin gene promoter; 88.
XX
OS Synthetic.
OS Pisum sativum.
XX
PN W0200056906-A1.
XX
PD 28-SEP-2000.
XX
PF 20-MAR-2000; 2000WO-IB00317.
XX
PR 22-MAR-1999; 99FR-0003635.
XX
PA (MER1-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-587667/55.
XX
PT Chimeric expression promoter for producing dicotyledonous and
PT monocotyledonous transgenic plants comprises a nucleic acid sequence
PT derived from a promoter of the pea plastocyanin gene
XX
PS Claim 2; Page 71; 83pp; English.
XX
CC The present sequence represents the chimeric promoter MP1111. The
CC promoter is derived from the petB promoter from pea plastocyanin gene,
CC by fusing the petB as-1 like and nos enhancer like elements to the
CC promoter MP1098 (comprising TATA and CAAT boxes of petB), and then
CC inserting a G box and fusing a fragment comprising a duplication of the
CC element as2 and as1. The petB promoter directs cell-specific but not
CC full light-regulated expression in transgenic tobacco plants. The
CC promoter is used to construct chimeric promoters of the invention. The
CC specification describes a chimeric expression promoter comprising a
CC promoter of the pea plastocyanin gene, or comprising a G box operably
CC or functionally linked upstream of a CAAT box, TATA box and transcription
CC initiation site. The chimeric promoters are used in expression vectors
CC for producing transgenic plants, such as dicotyledonous species,
CC e.g. potato, tobacco, cotton, lettuce, tomato, melon, cucumber, pea,
CC rape, beetroot, or sunflower, and monocotyledonous species, e.g. wheat,
CC barley, oat, rice, or corn.
XX
SQ Sequence 296 BP; 94 A; 74 C; 45 G; 83 T; 0 other;
Query Match 94.9%; Score 75; DB 21; Length 296;
Best Local Similarity 100.0%; Pred. No. 2.5e-18;

Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 CTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAAAGG 64
DB 13 CTGAGACTAGTGTGATGATGATCAAGATTGATGATCTCCACTGACGTAAAGG 72
QY 65 ATGACGATGCCACT 79
DB 73 ATGACGATGCCACT 87
RESULT 5
AAA96859
ID AAA96859 standard; DNA; 472 BP.
XX
AC AAA96859;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MP1169.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; 88.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN W0200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99FR-0003925.
XX
PA (MER1-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
PS Claim 5; Page 88; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 472 BP; 149 A; 92 C; 112 G; 119 T; 0 other;
Query Match 88.6%; Score 70; DB 21; Length 472;
Best Local Similarity 93.6%; Pred. No. 2.2e-16;
Matches 73; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 CAGCTGAGACTGATGATGATGATCAAGATTGATGATCTCCACTGACGTAA 60
DB 306 CTTGATTAAGACTGATGATGATGATCAAGATTGATGATCTCCACTGACGTAA 365
QY 61 AGGATGACGATGCCAC 78

KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
OS Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99PR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Claim 5; Page 86; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 462 BP; 148 A; 87 C; 111 G; 116 T; 0 other;
XX
Query Match 88.1%; Score 69.6; DB 21; Length 462;
Best Local Similarity 94.7%; Pred. No. 3e-16;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 TGCAGACAGTGTGATGATGATATCAAGATTGATGATATCTCCACTGACGTAAAG 62
DB 227 TGGTTACGACAGTGTGATGATGATATCAAGATTGATGATATCTCCACTGACGTAAAG 286
QY 63 GGATGACGACATGCCAC 78
DB 287 GGATGACGACATGCCAC 302
XX
RESULT 9
AAA96856 standard; DNA; 600 BP.
XX
AC AAA96856;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MPr1165.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.

XX
XX WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX
PR 29-MAR-1999; 99PR-0003925.
XX
PA (MERI-) MERISTEM THERAPEUTICS.
XX
PI Rance I, Gruber V, Theisen M;
XX
PI Rance I, Gruber V, Theisen M;
XX
DR WPI; 2000-647238/62.
XX
PT Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
PS Claim 5; Page 86-87; 91pp; English.
XX
CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
SQ Sequence 600 BP; 188 A; 111 C; 147 G; 154 T; 0 other;
XX
Query Match 88.1%; Score 69.6; DB 21; Length 600;
Best Local Similarity 94.7%; Pred. No. 3.3e-16;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 TGCAGACAGTGTGATGATGATATCAAGATTGATGATATCTCCACTGACGTAAAG 62
DB 227 TGGTTACGACAGTGTGATGATGATATCAAGATTGATGATATCTCCACTGACGTAAAG 286
QY 63 GGATGACGACATGCCAC 78
DB 287 GGATGACGACATGCCAC 302
XX
RESULT 10
AAA96855/c
ID AAA96855 standard; DNA; 392 BP.
XX
AC AAA96855;
XX
DT 19-FEB-2001 (first entry)
XX
DE Nucleotide sequence of chimeric expression promoter MPr1164.
XX
KM Promoter; intergenic region; Commelina yellow mottle virus;
KM chimeric expression promoter; plant vascular expression promoter;
KM plant green tissue expression promoter; Cassava vein mosaic virus;
KM transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
OS Chimeric - Cassava vein mosaic virus.
XX
PN WO200058485-A1.
XX
PD 05-OCT-2000.
XX
PF 29-MAR-2000; 2000WO-IB00370.
XX

```
PR 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
XX Claim 5; Page 86; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
XX Sequence 392 BP; 127 A; 80 C; 87 G; 98 T; 0 other;
XX
XX Query Match 87.3%; Score 69; DB 21; Length 392;
XX Best Local Similarity 100.0%; Pred. No. 4.8e-16;
XX Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 10 GACTAGTATGATGTGATATCAAGATTGATGATATCTCCACTGAGTAAGGATGAC 69
DB 302 GACTAGTATGATGTGATATCAAGATTGATGATATCTCCACTGAGTAAGGATGAC 243
XX
XX 70 GCATGCCAC 78
DB 242 GCATGCCAC 234
XX
XX
XX RESULT 11
XX ID AAA96857 standard; DNA; 604 BP.
XX
XX AAA96857;
XX
XX 19-FEB-2001 (first entry)
XX
XX Nucleotide sequence of chimeric expression promoter MPr1167.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
XX chimeric expression promoter; plant vascular expression promoter;
XX plant green tissue expression promoter; Cassava vein mosaic virus;
XX transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
XX Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
XX
XX 29-MAR-2000; 2000WO-IB00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
```

```
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
XX Claim 5; Page 87; 91pp; English.
XX
XX The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
XX Sequence 604 BP; 186 A; 116 C; 145 G; 157 T; 0 other;
XX
XX Query Match 86.3%; Score 68.2; DB 21; Length 604;
XX Best Local Similarity 95.9%; Pred. No. 1.1e-15;
XX Matches 70; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX
XX 6 TGCACTAGTATGATGTGATATCAAGATTGATGATATCTCCACTGAGTAAGGA 65
DB 376 TGCACTAGTATGATGTGATATCAAGATTGATGATATCTCCACTGAGTAAGGA 435
XX
XX 66 TGACGCATGCCAC 78
DB 436 TGACGCATGCCAC 448
XX
XX
XX RESULT 12
XX ID AAA96839 standard; DNA; 371 BP.
XX
XX AAA96839;
XX
XX 19-FEB-2001 (first entry)
XX
XX Nucleotide sequence of chimeric expression promoter MPr1146.
XX
XX Promoter; intergenic region; Commelina yellow mottle virus;
XX chimeric expression promoter; plant vascular expression promoter;
XX plant green tissue expression promoter; Cassava vein mosaic virus;
XX transgenic plant; chimera; ss.
XX
XX Chimeric - Commelina yellow mottle virus.
XX Chimeric - Cassava vein mosaic virus.
XX
XX WO200058485-A1.
XX
XX 05-OCT-2000.
XX
XX 29-MAR-2000; 2000WO-IB00370.
XX
XX 29-MAR-1999; 99FR-0003925.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Rance I, Gruber V, Theisen M;
XX
XX WPI; 2000-647238/62.
XX
XX Chimeric expression promoter for transgenic plant production, comprises
PT sequence from promoter comprising vascular expression region replaced
PT with sequence from promoter comprising green tissue expression region
XX
XX
XX Claim 5; Page 81; 91pp; English.
XX
```

CC The present sequence represents a chimeric promoter of the invention.
CC The specification describes chimeric expression promoters. These
CC chimeric promoters comprise a nucleic acid sequence which is derived
CC from a first plant promoter, in which a plant vascular expression
CC promoter region is replaced with a nucleic acid sequence derived from
CC a second plant promoter comprising a plant green tissue expression
CC promoter region. Preferably, the first plant promoter originates from
CC Commelina yellow mottle virus, and the second plant promoter originates
CC from the Cassava vein mosaic virus. Especially, the promoters are
CC derived from intergenic regions. The chimeric promoters are useful
CC for producing transgenic plants.
XX
XX Sequence 371 BP, 122 A; 68 C; 89 G; 92 T; 0 other;
XX
XX Query Match 86.1%; Score 68; DB 21; Length 371;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-15;
XX Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps
XX
XX 12 CTATGATTTGATGCTGATATCAGATTGATGTGATATCTTCACTGACGTAAAGGATGACGC 71
XX Db 71 CTATGATTTGATGCTGATATCAGATTGATGTGATATCTTCACTGACGTAAAGGATGACGC 130
XX
XX Oy 72 ATGCCACT 79
XX |||||
XX Db 131 ATGCCACT 138
XX
XX
XX RESULT 13
XX AAD03420/C
XX ID AAD03420 standard; DNA; 80 BP.
XX
XX AAD03420;
XX
XX DT 13-JUN-2001 (first entry)
XX
XX DE CaMV as-2/as-1 PCR primer #1 used for constructing MP1130 promoter.
XX
XX KW Wheat; glutenin; transgenic plant; PRHMG-Dx5; MP1130 promoter;
XX KW CaMV 35S promoter; as-2 motif; as-1 motif; PCR primer; ss.
XX
XX OS Cauliflower mosaic virus.
XX
XX OS Synthetic.
XX
XX PN WO200123593-A1.
XX
XX PD 05-APR-2001.
XX
XX PP 28-SEP-2000; 2000WO-1B01383.
XX
XX PR 30-SEP-1999; 99FR-0012373.
XX
XX PA (MERI-) MERISTEM THERAPEUTICS.
XX
XX PI Gruber V, Norre F, Theisen M;
XX DR MPI; 2001-245005/25.
XX
XX PT Chimeric expression promoters comprising nucleic acids from genes
XX PT encoding a high molecular weight wheat glutenin proteins, useful in the
XX PT production of transgenic plants -
XX
XX Example 3; Page 32; 120pp; English.
XX
XX The patent discloses chimeric expression promoters comprising nucleic
XX acid sequences derived from the promoters of genes encoding high
XX molecular weight wheat glutenin proteins. The chimeric promoters
XX are used in standard recombinant DNA techniques for the generation
XX of transgenic plants and for the production of polypeptides (e.g. marker
XX and therapeutic polypeptides).
XX The present DNA sequence is a PCR primer comprising the cauliflower
XX mosaic virus (CaMV) 35S promoter as-2 motifs (in duplicate) and as-1
XX motif, used in the construction of the wheat high molecular weight

CC	glutenin PrHMMG-Dx5 promoter-derived chimeric promoter Mprl130 and
CC	plasmids containing chimeric promoter sequences.
XX	
SQ	Sequence 80 BP; 21 A; 18 C; 16 G; 25 T; 0 other;
	Query Match 73.4%; Score 58; DB 22; Length 80;
	Best Local Similarity 100.0%; Pred. No. 3 8e-12;
	Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	15 GTGATTGATGTGATATCAAGATTGATGTGATATCTCCA CTGACGTAAAGCATGA CGCA 72 Db 58 GTGATTGATGTGATATCAAGATTGATGTGATATCTCCA CTGACGTAAAGCATGA CGCA 1
RESULT 14	
ID	AAD03396
XX	AAD03396 standard; DNA; 236 BP.
XX	AAD03396;
DT	13-JUN-2001 (first entry)
DE	Mprl133 promoter DNA derived from wheat glutenin MPrHMMG-Dx5 promoter.
XX	
KM	Wheat; glutenin; transgenic plant; Mprl133 promoter;
XX	CAMV 35S promoter; ds.
OS	Chimeric - Triticum aestivum.
XX	Chimeric - Cauliflower mosaic virus.
FH	Key
FT	enhancer
FT	Location/Qualifiers
FT	78..132
FT	/tag= a
FT	/note= "Aas2/Aas2/Aas1 box from CAMV 35S promoter"
FT	198
FT	/tag= b
FT	/note= "Transcription initiation site"
XX	
PN	WO200123593-A1.
PD	05-APR-2001.
PF	28-SEP-2000; 2000MO-IBO1383.
PR	30-SEP-1999; 99FR-0012373.
PA	(MERI-) MERISTEM THERAPEUTICS.
PI	Gruher V, Norre F, Theisen M;
DR	WPI; 2001-245005/25.
PT	Chimeric expression promoters comprising nucleic acids from genes
PT	encoding a high molecular weight wheat glutenin proteins, useful in the
PT	production of transgenic plants -
PS	
Claim 4; Page 116; 120pp; English.	
CC	The patent discloses chimeric expression promoters comprising nucleic
CC	acid sequences derived from the promoters of genes encoding high
CC	molecular weight wheat glutenin proteins. The chimeric promoters
CC	are used in standard recombinant DNA techniques for the generation
CC	of transgenic plants and the production of polypeptides (e.g. marker
CC	and therapeutic polypeptides).
CC	The present DNA sequence is Mprl133 promoter which is derived by
CC	deleting the sequence (comprising 2 prolamine-like boxes, 2 GA7A boxes,
CC	G-box and activating element) located upstream of the nucleotide -197
CC	of Mprl130 promoter (AAD03394). The Mprl130 promoter is derived from
CC	wheat high molecular weight glutenin PrHMMG-Dx5 promoter DNA (AAD03390)
CC	and contains elements from the cauliflower mosaic virus (CaMV) 35S
CC	promoter.
CC	Note: The present sequence is referred as SEQ ID NO: 7 throughout the
CC	specification. However this sequence is shown as SEQ ID NO: 35 in

CC the sequence listing.
XX
SQ Sequence 236 BP; 64 A; 70 C; 40 G; 62 T; 0 other;
Query Match 73.4%; Score 58; DB 22; Length 236;
Best Local Similarity 100.0%; Pred. No. 5.3e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 15 GTGATTGATGATATCAAGATTGATGATATCTCCACGTACGTAAGGATGACGCA 72
DB 75 GTGATTGATGATATCAAGATTGATGATATCTCCACGTACGTAAGGATGACGCA 132
RESULT 15
AAD03397
ID AAD03397 standard; DNA; 299 BP.
AC AAD03397;
XX
DT 13-JUN-2001 (first entry)
XX
DE Mpr1134 promoter DNA derived from wheat glutenin MprHMG-Dx5 promoter.
XX
KW wheat; glutenin; transgenic plant; Mpr1134 promoter;
KM CAMV 35S promoter; ds.
XX
OS Chimeric - Trifolium aestivum.
OS Chimeric - Cauliflower mosaic virus.
XX
XX
Key Location/Qualifiers
FH enhancer 20..57
FT /*tag= a
FT enhancer 141..195
FT /*tag= b
FT /*note= "As2/As2/As1 box from CAMV 35S promoter"
FT misc_signal 261
FT /*tag= C
FT /*note= "Transcription initiation site"
XX
PN NO200123593-A1.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-IB01383.
XX
PR 30-SEP-1999; 99FR-0012373.
XX
XX (MERI-) MERISTEM THERAPEUTICS.
XX
XX Gruber V, Norre F, Theisen M;
XX
XX WPI, 2001-245005/25.
XX
XX
XX Chimeric expression promoters comprising nucleic acids from genes
PT encoding a high molecular weight wheat glutenin proteins, useful in the
PT production of transgenic plants -
XX
XX
PS Claim 4; Page 116; 120pp; English.
XX
XX The patent discloses chimeric expression promoters comprising nucleic
CC acid sequences derived from the promoters of genes encoding high
CC molecular weight wheat glutenin proteins. The chimeric promoters
CC are used in standard recombinant DNA techniques for the generation
CC of transgenic plants and the production of polypeptides (e.g. marker
CC and therapeutic polypeptides).
CC The present DNA sequence is Mpr1134 promoter which is derived by
CC deleting the sequence (comprising 2 prolamine-like boxes, 2 GATA boxes
CC and G-like box) located upstream of the nucleotide -260 of Mpr1130
CC promoter (AAD03394). The Mpr1130 promoter is derived from wheat high
CC molecular weight glutenin PmHMG-Dx5 promoter DNA (AAD03390)
CC and contains elements from the cauliflower mosaic virus (CaMV) 35S
CC promoter.
CC Note: The present sequence is referred as SEQ ID NO: 8 throughout the

CC specification. However this sequence is shown as SEQ ID NO: 36 in
CC the sequence listing.
XX
SQ Sequence 299 BP; 78 A; 87 C; 49 G; 85 T; 0 other;
Query Match 73.4%; Score 58; DB 22; Length 299;
Best Local Similarity 100.0%; Pred. No. 5.7e-12;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 15 GTGATTGATGATATCAAGATTGATGATATCTCCACGTACGTAAGGATGACGCA 72
DB 138 GTGATTGATGATATCAAGATTGATGATATCTCCACGTACGTAAGGATGACGCA 195
Search completed: May 11, 2003, 03:04:04
Job time : 154.644 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:50:55 ; Search time 836.727 seconds
(without alignments)
2747.757 Million cell updates/sec

Title: US-09-963-803-14

Perfect score: 79
Sequence: 1 catgcgcgcagctagtgatt.....aaggatgacgcagccact 79

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_ov: *
5: gb_ov: *
6: gb_ph: *
7: gb_ph: *
8: gb_pl: *
9: gb_pl: *
10: gb_ro: *
11: gb_ro: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_or: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_stg: *
28: em_un: *
29: em_un: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rtd: *
36: em_hcg_mam: *
37: em_hcg_vrt: *
38: em_sy: *
39: em_higo_hum: *
40: em_higo_mus: *
41: em_higo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	79	100.0	79	6	AX036611 Sequence
2	79	100.0	79	6	AX036748 Sequence
3	75	94.9	259	6	AX036616 Sequence
4	75	94.9	296	6	AX036603 Sequence
5	70	88.6	472	6	AX036759 Sequence
6	70	88.6	541	6	AX036758 Sequence
7	69.6	88.1	393	6	AX036753 Sequence
8	69.6	88.1	462	6	AX036754 Sequence
9	69.6	88.1	600	6	AX036756 Sequence
10	69.6	88.1	9285	6	AX033047 Sequence
11	69.6	88.1	15077	6	AX033052 Sequence
12	69	87.3	392	6	AX036755 Sequence
13	68.2	86.3	604	6	AX036757 Sequence
14	68	86.1	371	6	AX036739 Sequence
15	58	73.4	80	6	AX103782 Sequence
16	58	73.4	236	6	AX103783 Sequence
17	58	73.4	299	6	AX103784 Sequence
18	58	73.4	332	6	AX103755 Sequence
19	58	73.4	472	6	AX103753 Sequence
20	55	69.6	80	6	AX103776 Sequence
21	46	58.2	63	6	AX036610 Sequence
22	46	58.2	63	6	AX036747 Sequence
23	46	58.2	280	6	AX036741 Sequence
24	46	58.2	301	6	AX036740 Sequence
25	46	58.2	398	6	AX036740 Sequence
26	39.2	49.6	63	6	AX103781 Sequence
27	39.2	49.6	219	6	AX103756 Sequence
28	39.2	49.6	282	6	AX103757 Sequence
29	39.2	49.6	315	6	AX103758 Sequence
30	39.2	49.6	381	6	AX103753 Sequence
31	39.2	49.6	455	6	AX103754 Sequence
32	39.2	49.6	505	6	AX103759 Sequence
33	39	49.4	174	8	AF434749 Zea may
34	39	49.4	189	8	AF434747 Zea may
35	39	49.4	197	8	AF434746 Zea may
36	39	49.4	199	8	AF434748 Zea may
37	39	49.4	199	8	AF434750 Zea may
38	39	49.4	199	8	AF434750 Zea may
39	39	49.4	240	6	AX033493 Sequence
40	39	49.4	309	6	AX044092 Sequence
41	39	49.4	309	6	AX040516 Sequence
42	39	49.4	331	6	BD001990 A transge
43	39	49.4	332	6	E01311 Cauliflower
44	39	49.4	333	6	I04847 Sequence 3
45	39	49.4	354	12	ARCAMPVR X04879 Canv Promot

ALIGNMENTS

RESULT 1
AX036611
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
FEATURES

AX036611
Sequence 17 from Patent PR2791358.
AX036611
AX036611.1 GI:11226206
79 bp
DNA
1linear
PAT 16-NOV-2000

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 79)
Rance, I., Theisen, M. and Gruber, V.
Patent: FR 2791358-A 17 29-SEP-2000;
MERISTEM THERAPEUTICS (FR)
Location/Qualifiers

```

source
1..79
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional building block S7-Directional building
block oligonucleotide for the construction of promoters by
1b-PCR"
BASE COUNT      22 a      15 c      20 g      22 t
ORIGIN
Query Match      100.0%; Score 79; DB 6; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.3e-17;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 60
    |||||||
Db 1 CATGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 60
    |||||||
QY 61 AGGATGACGCATGCCACT 79
    |||||||
Db 61 AGGATGACGCATGCCACT 79
    |||||||

RESULT 3
LOCUS      AX036616      259 bp      DNA      linear      PAT 16-NOV-2000
DEFINITION Sequence 22 from Patent FR2791358.
ACCESSION  AX036616
VERSION     AX036616.1 GI:11226211
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 259)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
PATENT: FR 2791358-A 22 29-SEP-2000;

source
1..79
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional building block S7-Directional building
block oligonucleotide for the construction of promoters by
1b-PCR"
BASE COUNT      22 a      15 c      20 g      22 t
ORIGIN
Query Match      100.0%; Score 79; DB 6; Length 79;
Best Local Similarity 100.0%; Pred. No. 1.3e-17;
Matches 79; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CATGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 60
    |||||||
Db 1 CATGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 60
    |||||||
QY 61 AGGATGACGCATGCCACT 79
    |||||||
Db 61 AGGATGACGCATGCCACT 79
    |||||||

RESULT 2
LOCUS      AX036748      79 bp      DNA      linear      PAT 16-NOV-2000
DEFINITION Sequence 14 from Patent WO0058485.
ACCESSION  AX036748
VERSION     AX036748.1 GI:11226257
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct
artificial sequences.
REFERENCE
1 (bases 1 to 79)
AUTHORS
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
PATENT: WO 0058485-A 14 05-OCT-2000;
JOURNAL
MEDISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
location/Qualifiers
1..79
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Directional desoxynucleotide building block S7"
BASE COUNT      22 a      15 c      20 g      22 t
ORIGIN

```

FEATURES	SOURCE	MERISTEM	THERAPEUTICS (FR)	Location/Qualifiers
1. .259				/organism="synthetic construct"
/db_xref="taxon:32630"				
1. .259				/note="Promoter MP1112 differs from MP1111 by a deletion of 35 bp containing the 'G' box and stretching from bp position -127 to position -89 and a deletion of two bp situated at positions -78 and -76 promoter MP1112"
BASE COUNT	82 a	67 c	38 g	72 t
ORIGIN				
Query Match	94.9%;	Score 75;	DB 6;	Length 259;
Best Local Similarity	100.0%;	Pred. No. 3,4e-16;		
Matches 75;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	5	CTGCAGACTAGTGGATGATGATATCAAGATGATGATATCTCCAGCTAAGAGG	64	
DB	13	CTGCAGACTAGTGGATGATGATATCAAGATGATGATATCTCCAGCTAAGAGG	72	
QY	65	ATGACGATGCCACT	79	
DB	73	ATGACGATGCCACT	87	
RESULT 4				
LOCUS	AX036603	296 bp	DNA	linear
DEFINITION	Sequence 9 from Patent FR2791358.			PAT 16-NOV-2000
ACCESSION	AX036603			
VERSION	AX036603.1	GI:11226198		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
JOURNAL				
MERISTEM				
FEATURES				
source				
1. .296				/organism="synthetic construct"
/db_xref="taxon:32630"				
1. .296				/note="Promoter MP1111 created by inserting at -99 bp position of MP1098, an 18 bp element containing a 'G' box and fusing a sequence of 58 bp (duplication of the element as2 and as1) promoter MP1111"
BASE COUNT	94 a	74 c	45 g	83 t
ORIGIN				
Query Match	94.9%;	Score 75;	DB 6;	Length 296;
Best Local Similarity	100.0%;	Pred. No. 3,4e-16;		
Matches 75;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	5	CTGCAGACTAGTGGATGATGATATCAAGATGATGATATCTCCAGCTAAGAGG	64	
DB	13	CTGCAGACTAGTGGATGATGATATCAAGATGATGATATCTCCAGCTAAGAGG	72	
QY	65	ATGACGATGCCACT	79	
DB	73	ATGACGATGCCACT	87	
RESULT 5				
LOCUS	AX036759	472 bp	DNA	linear
DEFINITION	Sequence 25 from Patent WO0058485.			PAT 16-NOV-2000
ACCESSION	AX036759			
VERSION	AX036759.1	GI:11226268		

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
promoter
BASE COUNT
ORIGIN

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 472)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 25 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .472
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1163"

149 a 92 c 112 g 119 t

Query Match 88.6%; Score 70; DB 6; Length 472;
Best Local Similarity 93.6%; Pred. No. 2.1e-14;
Matches 73; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CATCTGACAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 60
DB 306 CTGTGTTACGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 365
QY 61 AGGATGACGATGCCAC 78
DB 366 AGGATGACGATGCCAC 383

RESULT 6
LOCUS AX036758 541 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 24 from Patent WO0058485.
ACCESSION AX036758
VERSION AX036758.1 GI:11226267
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
promoter
BASE COUNT
ORIGIN

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 541)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 24 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .541
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1168"

169 a 104 c 130 g 138 t

Query Match 88.6%; Score 70; DB 6; Length 541;
Best Local Similarity 93.6%; Pred. No. 2.1e-14;
Matches 73; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CATCTGACAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 60
DB 306 CTGTGTTACGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTA 365
QY 61 AGGATGACGATGCCAC 78
DB 366 AGGATGACGATGCCAC 383

RESULT 7

AX036753
LOCUS AX036753 393 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 19 from Patent WO0058485.
ACCESSION AX036753
VERSION AX036753.1 GI:11226262
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
promoter
BASE COUNT
ORIGIN

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 393)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 19 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .393
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1162"

128 a 75 c 93 g 97 t

Query Match 88.1%; Score 69.6; DB 6; Length 393;
Best Local Similarity 94.7%; Pred. No. 3e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTAAG 62
DB 227 TGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTAAG 286
QY 63 GGATGACGATGCCAC 78
DB 287 GGATGACGATGCCAC 302

RESULT 8
LOCUS AX036754 462 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 20 from Patent WO0058485.
ACCESSION AX036754
VERSION AX036754.1 GI:11226263
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
promoter
BASE COUNT
ORIGIN

synthetic construct.
synthetic construct
artificial sequences.
1 (bases 1 to 462)
Rance, I., Theisen, M. and Gruber, V.
Chimeric expression promoters originating from commelina yellow
mottle virus and cassava vein mosaic virus
Patent: WO 0058485-A 20 05-OCT-2000;
MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
Location/Qualifiers
1. .462
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="promoter MP1163"

148 a 87 c 111 g 116 t

Query Match 88.1%; Score 69.6; DB 6; Length 462;
Best Local Similarity 94.7%; Pred. No. 2.9e-14;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 TGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTAAG 62
DB 227 TGCTGCAGACTAGTATGATGATATCAAGATTGATGATATCTCCACTGACGTAAG 286
QY 63 GGATGACGATGCCAC 78

Query Match	Best Local Similarity	Score	DB	Length
Query Match	Best Local Similarity	Score	DB	Length
Matches 72; Conservative	0; Mismatches 4; Indels 0; Gaps 0;			
LOCUS	AX036756	600 bp	DNA	linear
DEFINITION	Sequence 22 from Patent WO0058485.			PAT 16-NOV-2000
ACCESSION	AX036756			
VERSION	AX036756.1			
KEYWORDS				
SOURCE	synthetic construct.			
ORGANISM	artificial construct			
REFERENCE	artificial sequences.			
AUTHORS	1 (bases 1 to 600)			
TITLE	Rance, I., Theisen, M. and Gruber, V.			
JOURNAL	Chimeric expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus			
FEATURES	Patent: WO 0058485-A 22 05-OCT-2000;			
source	MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR) ; GRUBER VERONIQUE (FR)			
Location/Qualifiers				
1..600				
/organism="synthetic construct"				
/db_xref="taxon:32630"				
/note="promoter MPR1165"				
1..600				
Promoter	188 a 111 c 147 g 154 c			
BASE COUNT				
ORIGIN				
Query Match	Best Local Similarity	Score	DB	Length
Matches 72; Conservative	0; Mismatches 4; Indels 0; Gaps 0;			
LOCUS	AX093047	9285 bp	DNA	linear
DEFINITION	Sequence 52 from Patent WO0118192.			PAT 30-MAR-2001
ACCESSION	AX093047			
VERSION	AX093047.1			
KEYWORDS	GI:13509522			
SOURCE				
ORGANISM	synthetic construct.			
REFERENCE	synthetic construct			
AUTHORS	artificial sequences.			
TITLE	1 (bases 1 to 9285)			
JOURNAL	Gruber, V. and Comeau, D.			
FEATURES	Synthetic vectors, transgenic plants containing them, and methods for obtaining them			
source	Patent: WO 0118192-A 52 15-MAR-2001;			
Location/Qualifiers	MERISTEM THERAPEUTICS (FR)			
1..9285				
/organism="synthetic construct"				
/db_xref="taxon:32630"				
/note="PMR11336"				
1				
/note="PMR11336 results from the insertion into PMR1196 of the promoter MPR165 isolated from plasmid PMR1322 as described in PCT patent application PCT/IB00/00370"				
BASE COUNT	2440 a 2252 c 2506 g 2087 t			
ORIGIN				
Query Match	Best Local Similarity	Score	DB	Length
Matches 72; Conservative	0; Mismatches 4; Indels 0; Gaps 0;			
LOCUS	AX036756	600 bp	DNA	linear
DEFINITION	Sequence 22 from Patent WO0058485.			PAT 16-NOV-2000
ACCESSION	AX036756			
VERSION	AX036756.1			
KEYWORDS				
SOURCE	synthetic construct.			
ORGANISM	artificial construct			
REFERENCE	artificial sequences.			
AUTHORS	1 (bases 1 to 600)			
TITLE	Rance, I., Theisen, M. and Gruber, V.			
JOURNAL	Chimeric expression promoters originating from commelina yellow mottle virus and cassava vein mosaic virus			
FEATURES	Patent: WO 0058485-A 22 05-OCT-2000;			
source	MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR) ; GRUBER VERONIQUE (FR)			
Location/Qualifiers				
1..600				
/organism="synthetic construct"				
/db_xref="taxon:32630"				
/note="promoter MPR1165"				
1..600				
Promoter	188 a 111 c 147 g 154 c			
BASE COUNT				
ORIGIN				

Matches	72:	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;
Qy	3	TGCTGCAGACTAGTATGATGTGATATCAAGATTGATGATATCTCCACTGACGTAAAG	62						
Db	6016	TGGTTACACACTAGTAGTATGATGTGATATCAAGATTGATGATATCTCCACTGACGTAAAG	6075						
Qy	63	GGATGACCATGCGCAC	78						
Db	6076	GGATGACCATGCGCAC	6091						
RESULT 11									
LOCUS	AX093052		15077 bp	DNA		linear		PAT 30-MAR-2001	
DEFINITION	Sequence 57 from Patent WO0118192.								
ACCESSION	AX093052								
VERSION	AX093052.1	GI:13509527							
KEYWORDS									
SOURCE									
ORGANISM		synthetic construct.							
REFERENCE		synthetic construct							
AUTHORS		artificial sequences.							
TITLE		1 (bases 1 to 15077)							
JOURNAL		Gruber,V. and Comeau,D.							
		Synthetic vectors, transgenic plants containing them, and methods							
		for obtaining them							
		Patent: WO 0118192-A 57 15-MAR-2001;							
		MERISTEM THERAPEUTICS (FR)							
FEATURES									
source		location/Qualifiers							
		1..15077							
		/organism="synthetic construct"							
		/db_xref="taxon:32630"							
		/note="pMRT1342"							
		1							
		/note="pMRT1342 results from the replacement of the							
		expression cassette ep35S-gus-polyA35S from pMRT1335 by							
		the expression cassette L5-gus-polyA35S isolated from							
		pMRT1336"							
BASE COUNT	3672 a	3892 c	4225 g	3288 t					
ORIGIN									
Query Match		88.1%; Score 69.6; DB 6; Length 15077;							
Best Local Similarity	94.7%; Pred. No. 2,7e-14;								
Matches	72; Conservative	0; Mismatches	4; Indels	0; Gaps	0;				
Qy	3	TGCTGCAGACTAGTATGATGTGATATCAAGATTGATGATATCTCCACTGACGTAAAG	62						
Db	7055	TGGTTACACACTAGTAGTATGATGTGATATCAAGATTGATGATATCTCCACTGACGTAAAG	7114						
Qy	63	GGATGACCATGCGCAC	78						
Db	7115	GGATGACCATGCGCAC	7130						
RESULT 12									
LOCUS	AX036755/c		392 bp	DNA		linear		PAT 16-NOV-2000	
DEFINITION	Sequence 21 from Patent WO0058485.								
ACCESSION	AX036755								
VERSION	AX036755.1	GI:11226264							
KEYWORDS									
SOURCE									
ORGANISM		synthetic construct.							
REFERENCE		synthetic construct							
AUTHORS		artificial sequences.							
TITLE		1 (bases 1 to 392)							
JOURNAL		Rance,I., Theisen,M. and Gruber,V.							
		Chimeric expression promoters originating from commelina yellow							
		mottle virus and cassava vein mosaic virus							
		Patent: WO 0058485-A 21 05-OCT-2000;							
		MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)							
		GRUBER VERONIQUE (FR)							
		location/Qualifiers							
		1..392							
		/organism="synthetic construct"							

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:44:10 ; Search time 31.6792 Seconds

(without alignments)
764.775 Million cell updates/sec

Title: US-09-963-803-14

Perfect score: 79
Sequence: 1 catgctgcagactgactgatt.....aaggatgacatgactgact 79

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_MA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	49.4	439	1	US-08-247-809A-3
2	39	49.4	439	2	US-08-711-728-3
3	39	49.4	446	1	US-08-764-100-23
4	39	49.4	532	3	US-09-042-426-1
5	39	49.4	532	4	US-09-281-238-1
6	39	49.4	532	4	US-09-330-760-1
7	39	49.4	532	4	US-09-328-473-1
8	39	49.4	532	4	US-09-330-737-1
9	39	49.4	532	4	US-09-329-169-1
10	39	49.4	532	4	US-09-330-714A-1
11	39	49.4	532	4	US-09-328-826-1
12	39	49.4	560	3	US-09-042-426-5
13	39	49.4	560	4	US-09-291-238-5
14	39	49.4	560	4	US-09-330-760-5
15	39	49.4	560	4	US-09-328-473-5
16	39	49.4	560	4	US-09-330-737-5
17	39	49.4	560	4	US-09-329-169-5
18	39	49.4	560	4	US-09-330-714A-5
19	39	49.4	560	4	US-09-328-826-5
20	39	49.4	661	4	US-09-027-998A-33
21	39	49.4	793	4	US-08-371-764-1
22	39	49.4	793	4	US-08-897-736-1
23	39	49.4	831	1	US-08-450-834-5
24	39	49.4	950	6	5177308-3
25	39	49.4	978	1	US-08-446-486-31
26	39	49.4	978	1	US-08-463-308-31
27	39	49.4	979	1	US-08-446-486-30

28	39	49.4	979	1	US-08-463-308-30	Sequence 30, Appl
29	39	49.4	980	6	5254799-30	Patent No. 5254799
30	39	49.4	1030	4	US-07-936-163-46	Sequence 46, Appl
31	39	49.4	1030	4	US-08-729-601A-43	Sequence 43, Appl
32	39	49.4	1034	4	US-09-363-970-35	Sequence 35, Appl
33	39	49.4	1138	4	US-09-011-151-8	Sequence 8, Appl
34	39	49.4	1138	4	US-09-011-151-9	Sequence 9, Appl
35	39	49.4	1136	4	US-08-729-601A-46	Sequence 46, Appl
36	39	49.4	1279	4	US-09-185-244-2	Sequence 2, Appl
37	39	49.4	1279	4	US-09-471-913-6	Sequence 6, Appl
38	39	49.4	1303	3	US-08-894-440-2	Sequence 2, Appl
39	39	49.4	1303	4	US-09-458-093-2	Sequence 2, Appl
40	39	49.4	1651	3	US-09-065-999-5	Sequence 5, Appl
41	39	49.4	1651	3	US-09-065-999-6	Sequence 6, Appl
42	39	49.4	1722	1	US-08-247-809A-5	Sequence 5, Appl
43	39	49.4	1722	2	US-08-711-728-5	Sequence 5, Appl
44	39	49.4	1829	1	US-07-966-187-17	Sequence 17, Appl
45	39	49.4	1863	1	US-08-525-507-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-08-247-809A-3
Sequence 3, Application US/08247809A
Patent No. 5569823
GENERAL INFORMATION:
APPLICANT: Peter H. Schreier, Klaus Stenzel, Gunter Adam;
APPLICANT: Edgar Maiss
TITLE OF INVENTION: DEOXYRIBONUCLEIC ACIDS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSER: SPRUNG HORN KRAMER & WOODS
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
MEDIUM TYPE: storage
COMPUTER: NEC Powermate 1 Plus
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/247,809A
FILING DATE: May 23, 1994
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P 43 178 45.6 (Germany)
FILING DATE: May 28, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9049-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 439 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-247-809A-3

Query Match 49.4%; Score 39; DB 1; Length 439;
Best Local Similarity 100.0%; Pred. No. 6-2e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 338 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 376

RESULT 2

US-08-711-728-3
Sequence 3, Application US/08711728
Patent No. 5973135
GENERAL INFORMATION:
APPLICANT: Peter H. Schreier, Klaus Stenzel, Gunter Adam;
APPLICANT: Edgar Maiss
TITLE OF INVENTION: DEOXYRIBONUCLEIC ACIDS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: SPRUNG HORN KRAMER & WOODS
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB
COMPUTER: NEC Powermate 1 Plus
OPERATING SYSTEM: DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/711.728
FILING DATE: 03-SEPT-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,809
FILING DATE: 23-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 43178456
FILING DATE: 28-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9049.1-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 439 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-711-728-3

Query Match 49.4%; Score 39; DB 2; Length 439;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 338 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 376

RESULT 3

US-08-764-100-23
Sequence 2, Application US/08764100
Patent No. 5773700
GENERAL INFORMATION:
APPLICANT: van Grinsven J., Martinus O.
APPLICANT: De Haan, Petrus T.
APPLICANT: Gielens L., Johannes J.
APPLICANT: Peters, Dirk
APPLICANT: Goldbach, Robert W.

TITLE OF INVENTION: Improvements in or Relating to Organic
TITLE OF INVENTION: Compounds
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sandoz Agro, Inc
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:
APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 57337001s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 446 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-23

Query Match 49.4%; Score 39; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 4

US-09-042-426-1
Sequence 1, Application US/09042426
Patent No. 6114608
GENERAL INFORMATION:
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6114608artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,426
FILING DATE: March 13, 1998

CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-042-426-1

Query Match 49.4%; Score 39; DB 3; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 5

US-09-291-238-1
Sequence 1, Application US/09291238
Patent No. 6222104

GENERAL INFORMATION:
APPLICANT: Irvan J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6222104artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/291.238
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-291-238-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 6

US-09-330-760-1
Sequence 1, Application US/09330760
Patent No. 6229075

GENERAL INFORMATION:
APPLICANT: Irvan J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6229075artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/330,760
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-330-760-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
DB 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 7

US-09-328-473-1
Sequence 1, Application US/09328473
Patent No. 6232533

GENERAL INFORMATION:
APPLICANT: Irvn J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 623253artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/328,473
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-328-473-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
|||||
Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 8
US-09-330-737-1
Sequence 1, Application US/09330737
Patent No. 6232534
GENERAL INFORMATION:
APPLICANT: Irvn J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 623253artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/330,737
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-330-737-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
|||||
Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 9
US-09-329-169-1
Sequence 1, Application US/09329169
Patent No. 6329575
GENERAL INFORMATION:
APPLICANT: Irvn J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 6329575artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/329,169
FILING DATE: 09-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 358 Promoter
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-329-169-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 10
US-09-330-714A-1
Sequence 1, Application US/09330714A
Patent No. 6342660
GENERAL INFORMATION:
APPLICANT: Irvan J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
Thuringiensis Gene
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6342660artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/330.714A
FILING DATE: 11-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042.426
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 358 Promoter
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-330-714A-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72

Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 11
US-09-328-826-1
Sequence 1, Application US/09328826
Patent No. 6399860
GENERAL INFORMATION:
APPLICANT: Irvan J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6399860artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/328.826
FILING DATE: 09-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/042.426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 358 Promoter
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-328-826-1

Query Match 49.4%; Score 39; DB 4; Length 532;
Best Local Similarity 100.0%; Pred. No. 6.6e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
Db 274 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 312

RESULT 12
US-09-042-426-5
Sequence 5, Application US/09042426
Patent No. 6114608
GENERAL INFORMATION:
APPLICANT: Irvan J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6114608artis Corporation
STREET: 564 Morris Avenue
CITY: Summit

STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,426
FILING DATE: March 13, 1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-042-426-5

Query Match 49.4%; Score 39; DB 3; Length 560;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
Db 322 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 360

RESULT 13
US-09-291-238-5
Sequence 5, Application US/09291238
Patent No. 6222104
GENERAL INFORMATION:
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6222104artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/291,238
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-291-238-5

Query Match 49.4%; Score 39; DB 4; Length 560;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
Db 322 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 360

RESULT 14
US-09-330-760-5
Sequence 5, Application US/09330760
Patent No. 6229075
GENERAL INFORMATION:
APPLICANT: Irvin J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
TITLE OF INVENTION: DNA Construct Containing Bacillus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6229075artis Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/330,760
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/042,426
FILING DATE: March 13, 1998
ATTORNEY/AGENT INFORMATION:
NAME: Hoxie, Thomas
REGISTRATION NUMBER: 32,993
REFERENCE/DOCKET NUMBER: 135/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: 35S Promoter
US-09-330-760-5

Query Match 49.4%; Score 39; DB 4; Length 560;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
 |||||
 DB 322 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 360

RESULT 15

US-09-328-473-5
 ; Sequence 5, Application US/09328473
 ; Patent No. 6232533
 ; GENERAL INFORMATION:
 ; APPLICANT: Irvan J. Mettler, Paul S. Dietrich, Ralph Sinibaldi
 ; TITLE OF INVENTION: DNA Construct Containing Bacillus
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 6232533artis Corporation
 ; STREET: 564 Morris Avenue
 ; CITY: Summit
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07901
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/328,473
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 09/042,426
 ; FILING DATE: March 13, 1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hoxie, Thomas
 ; REGISTRATION NUMBER: 32,993
 ; REFERENCE/DOCKET NUMBER: 135/1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (919) 541-8614
 ; TELEFAX: (919) 541-8689
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 560 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; IMMEDIATE SOURCE:
 ; CLONE: 358 Promoter
 ; US-09-328-473-5

Query Match 49.4%; Score 39; DB 4; Length 560;
 Best Local Similarity 100.0%; Pred. No. 6.7e-06;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
 |||||
 DB 322 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 360

Search completed: May 11, 2003, 03:07:42;
 Job time : 33.6792 secs

Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 25
LENGTH: 472
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1169
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(472)
OTHER INFORMATION:
US-09-963-803-25

Query Match 88.6%; Score 70; DB 9; Length 472;
Best Local Similarity 93.6%; Pred. No. 7.5e-16;
Matches 73; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CATCTCGACAGTATGATGATGATATCAAGATTGATGATATCTCCAGTACGTA 60
DB 306 CTGTGACAGTATGATGATGATGATATCAAGATTGATGATATCTCCAGTACGTA 365

OY 61 AGGATGACGATGCCAC 78
DB 366 AGGATGACGATGCCAC 383

RESULT 3
US-09-963-803-24
Sequence 24, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 24
LENGTH: 541
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1168
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(541)
OTHER INFORMATION:
US-09-963-803-24

Query Match 88.6%; Score 70; DB 9; Length 541;
Best Local Similarity 93.6%; Pred. No. 7.8e-16;
Matches 73; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 CATCTCGACAGTATGATGATGATATCAAGATTGATGATATCTCCAGTACGTA 60

DB 306 CTGTGACAGTATGATGATGATATCAAGATTGATGATATCTCCAGTACGTA 365
OY 61 AGGATGACGATGCCAC 78
DB 366 AGGATGACGATGCCAC 383

RESULT 4
US-09-963-803-19
Sequence 19, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 19
LENGTH: 393
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1162
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(393)
OTHER INFORMATION:
US-09-963-803-19

Query Match 88.1%; Score 69.6; DB 9; Length 393;
Best Local Similarity 94.7%; Pred. No. 9.8e-16;
Matches 72; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 TGCTGACAGTATGATGATGATATCAAGATTGATGATATCTCCAGTACGTA 62
DB 227 TGCTGACAGTATGATGATGATATCAAGATTGATGATATCTCCAGTACGTA 286

OY 63 GGATGACGATGCCAC 78
DB 287 GGATGACGATGCCAC 302

RESULT 5
US-09-963-803-20
Sequence 20, Application US/09963803
Publication No. US20030028922A1
GENERAL INFORMATION:
APPLICANT: Meristem Therapeutics
TITLE OF INVENTION: Chimeric expression promoters originating from commelina yellow
FILE REFERENCE: 184332042
CURRENT APPLICATION NUMBER: US/09/963,803
CURRENT FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: FR 99/03925
PRIOR FILING DATE: 1999-03-29
PRIOR APPLICATION NUMBER: PCT IB00/00370
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20
LENGTH: 462
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: promoter MP1163

Query Match 88.6%; Score 70; DB 9; Length 462;
Best Local Similarity 93.6%; Pred. No. 7.8e-16;
Matches 72; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match	73.4%	Score 58	DB 10	Length 80
Best Local Similarity	100.0%	Pred. No. 9.3e-12		
Matches 58	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	15	GTGATTGATGTCATATCAAGATTGATGTCATATCTCCACTGACGTAAAGGATGACGCA	72	
Db	58	GTGATTGATGTCATATCAAGATTGATGTCATATCTCCACTGACGTAAAGGATGACGCA	1	

Query Match	73.4%	Score 58	DB 10	Length 299
Best Local Similarity	100.0%	Pred. NO.	1.5e-11	
Matches 58	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY 15 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 72
 DB 138 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 195

RESULT 13

US-09-870-375-7
 ; Sequence 7, Application US/09870375
 ; Patent No. US20020083486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MERISTEM THERAPEUTICS
 ; TITLE OF INVENTION: SYNTHETIC AND CHIMERIC PROMOTERS, EXPRESSION CASSETTES,
 ; TITLE OF INVENTION: PLASMIDS, VECTORS, TRANSGENIC PLANTS ET SEEDS INCLUDING
 ; FILE REFERENCE: PTHMWG1
 ; CURRENT APPLICATION NUMBER: US/09/870,375
 ; CURRENT FILING DATE: 2001-05-30
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 332
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MP1135
 ; NAME/KEY: misc feature
 ; LOCATION: (21)..(28)
 ; OTHER INFORMATION: G-like box
 ; NAME/KEY: enhancer
 ; LOCATION: (53)..(90)
 ; OTHER INFORMATION: Enhancer box
 ; NAME/KEY: misc feature
 ; LOCATION: (174)..(228)
 ; OTHER INFORMATION: As2/As2/As1 box
 ; NAME/KEY: TATA signal
 ; LOCATION: (264)..(270)
 ; OTHER INFORMATION: TATA box
 ; NAME/KEY: misc feature
 ; LOCATION: (294)
 ; OTHER INFORMATION: Transcription Initiation Site
 ; US-09-870-375-7

Query Match 73.4%; Score 58; DB 10; Length 332;
 Best Local Similarity 100.0%; Pred. No. 1.5e-11;

Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 72
 DB 171 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 228

RESULT 14

US-09-870-375-5
 ; Sequence 5, Application US/09870375
 ; Patent No. US20020083486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MERISTEM THERAPEUTICS
 ; TITLE OF INVENTION: SYNTHETIC AND CHIMERIC PROMOTERS, EXPRESSION CASSETTES,
 ; TITLE OF INVENTION: PLASMIDS, VECTORS, TRANSGENIC PLANTS ET SEEDS INCLUDING
 ; FILE REFERENCE: PTHMWG1
 ; CURRENT APPLICATION NUMBER: US/09/870,375
 ; CURRENT FILING DATE: 2001-05-30
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 472
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MP1130
 ; OTHER INFORMATION: promoter

NAME/KEY: misc feature
 LOCATION: (22)..(29)
 OTHER INFORMATION: Proline-1-like box
 NAME/KEY: misc feature
 LOCATION: (70)..(73)
 OTHER INFORMATION: GATA box
 NAME/KEY: misc feature
 LOCATION: (87)..(90)
 OTHER INFORMATION: GATA box
 NAME/KEY: misc feature
 LOCATION: (127)..(133)
 OTHER INFORMATION: Proline-1-like box
 NAME/KEY: misc feature
 LOCATION: (161)..(168)
 OTHER INFORMATION: G-like box
 NAME/KEY: enhancer
 LOCATION: (193)..(230)
 OTHER INFORMATION: Enhancer box
 NAME/KEY: misc feature
 LOCATION: (314)..(368)
 OTHER INFORMATION: As2/As2/As1 box
 NAME/KEY: TATA signal
 LOCATION: (404)..(410)
 OTHER INFORMATION: TATA box
 NAME/KEY: misc feature
 LOCATION: (434)
 OTHER INFORMATION: Transcription Initiation Site
 ; US-09-870-375-5

Query Match 73.4%; Score 58; DB 10; Length 472;
 Best Local Similarity 100.0%; Pred. No. 1.7e-11;
 Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 72
 DB 311 GTGATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 368

RESULT 15

US-09-870-375-28
 ; Sequence 28, Application US/09870375
 ; Patent No. US20020083486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MERISTEM THERAPEUTICS
 ; TITLE OF INVENTION: SYNTHETIC AND CHIMERIC PROMOTERS, EXPRESSION CASSETTES,
 ; TITLE OF INVENTION: PLASMIDS, VECTORS, TRANSGENIC PLANTS ET SEEDS INCLUDING
 ; FILE REFERENCE: PTHMWG1
 ; CURRENT APPLICATION NUMBER: US/09/870,375
 ; CURRENT FILING DATE: 2001-05-30
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 28
 ; LENGTH: 80
 ; TYPE: DNA
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial
 ; OTHER INFORMATION: Sequence:Oligodeoxynucleotide
 ; US-09-870-375-28

Query Match 69.6%; Score 55; DB 10; Length 80;
 Best Local Similarity 100.0%; Pred. No. 1.1e-10;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 ATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 72
 DB 1 ATTGATGATATCAAGATTGATGATATCTCCACTGACGTAAGGATGACGCA 55

Search completed: May 11, 2003, 06:11:43
 Job time : 79.208 secs

MERISTEM THERAPEUTICS (FR) ; RANCE IANN (FR) ; THEISEN MANFRED (FR)
; GRUBER VERONIQUE (FR)
FEATURES
SOURCE Location/Qualifiers
1..34
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Guide desoxynucleotide building block G1"

BASE COUNT 8 a 9 c 7 g 10 t
ORIGIN

Query Match 100.0%; Score 34; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 9.1e-05;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GACTCCTACTTATGATCGTACTGAGACA 34
Db 1 GACTCCTACTTATGATCGTACTGAGACA 34

RESULT 2
AC024376 66624 bp DNA linear HTG 05-JAN-2001
LOCUS
DEFINITION Homo sapiens clone RP11-610N12, LOW-PASS SEQUENCE SAMPLING.
AC024376
VERSION AC024376.3 GI:12039551
KEYWORDS HTG; HTGS PHASED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 66624)
TITLE Birren,B., Lincon,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens, clone RP11-610N12
REFERENCE
2 (bases 1 to 66624)
TITLE
AUTHORS Birren,B., Lincon,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castele,A.,
Chenopel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domingo,M., Doyle,M.,
Fenebor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galegan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lenoczky,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., MacDonald,P., Marguis,N., McCarthy,M.,
McBwan,P., McGurt,A., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severey,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirelli,A.,
Traversan,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 5, 2001 this sequence version replaced gi:1767813.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6257
Center clone name: 610_N_12
* NOTE: This record contains 82 individual

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
711 810: contig of 710 bp in length
811 1530: contig of 720 bp in length
1531 1630: gap of 100 bp
1631 2328: contig of 698 bp in length
2328 2428: gap of 100 bp
2429 3164: contig of 736 bp in length
3165 3264: gap of 100 bp
3265 3971: contig of 707 bp in length
3972 4071: gap of 100 bp
4072 4770: contig of 699 bp in length
4771 4870: gap of 100 bp
4871 5593: contig of 723 bp in length
5594 5693: gap of 100 bp
5694 6416: contig of 723 bp in length
6417 6516: gap of 100 bp
6517 7230: contig of 714 bp in length
7231 7330: gap of 100 bp
7331 8041: contig of 711 bp in length
8042 8141: gap of 100 bp
8142 8862: contig of 721 bp in length
8863 8962: gap of 100 bp
8963 9693: contig of 731 bp in length
9694 9793: gap of 100 bp
9794 10506: contig of 713 bp in length
10507 10606: gap of 100 bp
10607 11318: contig of 712 bp in length
11319 11418: gap of 100 bp
11419 12132: contig of 714 bp in length
12133 12232: gap of 100 bp
12233 12950: contig of 718 bp in length
12951 13050: gap of 100 bp
13051 13772: contig of 722 bp in length
13773 13872: gap of 100 bp
13873 14584: contig of 712 bp in length
14585 14684: gap of 100 bp
14685 15399: contig of 715 bp in length
15400 15499: gap of 100 bp
15500 16197: contig of 698 bp in length
16198 16297: gap of 100 bp
16298 17002: contig of 705 bp in length
17003 17102: gap of 100 bp
17103 17825: contig of 723 bp in length
17826 17925: gap of 100 bp
17926 18645: contig of 720 bp in length
18646 18745: gap of 100 bp
18746 19459: contig of 714 bp in length
19460 19559: gap of 100 bp
19560 20279: contig of 720 bp in length
20280 20379: gap of 100 bp
20380 21081: contig of 702 bp in length
21082 21181: gap of 100 bp
21182 21887: contig of 706 bp in length
21888 21987: gap of 100 bp
21988 22699: contig of 712 bp in length
22700 22799: gap of 100 bp
22800 23519: contig of 720 bp in length
23520 23619: gap of 100 bp
23620 24366: contig of 747 bp in length
24367 24466: gap of 100 bp
24467 25185: contig of 719 bp in length
25186 25285: gap of 100 bp
25286 25989: contig of 704 bp in length

```

* 25990 26089: gap of 100 bp
* 26090 26791: contig of 702 bp in length
* 26792 26891: gap of 100 bp
* 26892 27595: contig of 704 bp in length
* 27596 27695: gap of 100 bp
* 27696 28422: contig of 727 bp in length
* 28423 28522: gap of 100 bp
* 28523 29222: contig of 700 bp in length
* 29223 29322: gap of 100 bp
* 29323 30040: contig of 718 bp in length
* 30041 30140: gap of 100 bp
* 30141 30860: contig of 720 bp in length
* 30861 30960: gap of 100 bp
* 30961 31694: contig of 734 bp in length
* 31695 31794: gap of 100 bp
* 31795 32514: contig of 720 bp in length
* 32515 32614: gap of 100 bp
* 32615 33313: contig of 699 bp in length
* 33314 33413: gap of 100 bp
* 33414 34127: contig of 714 bp in length
* 34128 34227: gap of 100 bp
* 34228 34944: contig of 717 bp in length
* 34945 35044: gap of 100 bp
* 35045 35753: contig of 709 bp in length
* 35754 35853: gap of 100 bp
* 35854 36566: contig of 713 bp in length
* 36567 36666: gap of 100 bp
* 36667 37401: contig of 735 bp in length
* 37402 37501: gap of 100 bp
* 37502 38204: contig of 703 bp in length
* 38205 38304: gap of 100 bp
* 38305 39011: contig of 707 bp in length
* 39012 39111: gap of 100 bp
* 39112 39824: contig of 713 bp in length
* 39825 39924: gap of 100 bp
* 39925 40647: contig of 723 bp in length
* 40648 40747: gap of 100 bp
* 40748 41463: contig of 716 bp in length
* 41464 41563: gap of 100 bp
* 41564 42286: contig of 723 bp in length
* 42287 42386: gap of 100 bp
* 42387 43081: contig of 695 bp in length
* 43082 43181: gap of 100 bp
* 43182 43881: contig of 700 bp in length
* 43882 43981: gap of 100 bp
* 43982 44714: contig of 733 bp in length
* 44715 44814: gap of 100 bp
* 44815 45524: contig of 710 bp in length
* 45525 45624: gap of 100 bp
* 45625 46313: contig of 689 bp in length
* 46314 46413: gap of 100 bp
* 46414 47144: contig of 731 bp in length
* 47145 47244: gap of 100 bp
* 47245 47974: contig of 730 bp in length
* 47975 48074: gap of 100 bp
* 48075 48793: contig of 719 bp in length
* 48794 48893: gap of 100 bp
* 48894 49612: contig of 719 bp in length
* 49613 49712: gap of 100 bp
* 49713 50414: contig of 702 bp in length
* 50415 50514: gap of 100 bp
* 50515 51120: contig of 606 bp in length
* 51121 51220: gap of 100 bp
* 51221 51934: contig of 714 bp in length
* 51935 52034: gap of 100 bp
* 52035 52751: contig of 717 bp in length
* 52752 52851: gap of 100 bp
* 52852 53560: contig of 709 bp in length
* 53561 53660: gap of 100 bp
* 53661 54381: contig of 721 bp in length
* 54382 54481: gap of 100 bp
* 54482 55194: contig of 713 bp in length
* 55195 55294: gap of 100 bp

```

```

* 55295 56008: contig of 714 bp in length
* 56009 56108: gap of 100 bp

Query Match
Best Local Similarity 63.5%; Score 21.6; DB 2; Length 66624;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 28524 GATCCCTACTTATGATCGTACTGTG 29
|||||
|GATCCCTACTTATGATCGTACTGTG 29
|||||

RESULT 3
AC091523 235505 bp DNA linear ROD 19-FEB-2002
LOCUS Mus musculus chromosome 1 clone rp23-77a8, complete sequence.
DEFINITION AC091523
VERSION AC091523.22 GI:18702401
KEYWORDS HTG.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 235505)
AUTHORS Deschamps, S., Chambliss, D. and Roe, B.A.
TITLE Mus musculus Chromosome 1 BAC Clone rp23-77a8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 235505)
AUTHORS Deschamps, S., Chambliss, D. and Roe, B.A.
TITLE Direct Submision
JOURNAL Submitted (29-FEB-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

COMMENT On Feb 19, 2002 this sequence version replaced gi:16604017.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----

FEATURES
source Location/Qualifiers
1..235505
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="1"
/clone="rp23-77a8"
/clone_1ib="RPCL Mouse BAC Library 23"

BASE COUNT 69731 a 50058 c 49581 g 66135 t
ORIGIN

Query Match
Best Local Similarity 80.6%; Score 21.4; DB 10; Length 235505;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 154904 CTCCTTACTTATGATCGTACTGTGAC 33
|||||
|CTCCTTACTTATGATCGTACTGTGAC 33
|||||

RESULT 4
AC123002 137875 bp DNA linear HTG 24-JUL-2002
LOCUS Rattus norvegicus clone CH230-22H7, *** SEQUENCING IN PROGRESS ***
DEFINITION AC123002
VERSION AC123002.2 GI:21909223
KEYWORDS HTG; HTGS_PHAISEL.
SOURCE Rattus norvegicus.

```

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 137875)
Munzy,D.M., Adams,C., Adlo-Ochola,B., Ali-osman,F.R., Allen,C.,
Albrooke,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbata,J., Benton,J., Bimege,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Briteva,M., Brown,E., Brown,M., Bryant,N.P.,
Burch,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dachorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Einhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homesi,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loussaged,H.,
Lozade,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,U.,
Mabeswari,M., Mapua,P., Martin,R., Matindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Mosier,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okumaru,G.,
Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivera,M., Rojas,A., Rojibokan,I., Roife,M., Ruiz,S., Saverly,G.,
Schreier,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swalek,A., Tabor,P., Tameria,A., Tamerias,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umanu,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Consensus quality: 97350 bases at least Q30
Consensus quality: 101050 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a "working draft" sequence. It currently
consists of 63 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1084: contig of 1084 bp in length
1085 1184: gap of unknown length
1185 2985: contig of 1801 bp in length
2986 3085: gap of unknown length
3086 4594: contig of 1509 bp in length
4595 4694: gap of unknown length
4695 6070: contig of 1376 bp in length
6071 6170: gap of unknown length
6171 7704: contig of 1534 bp in length
7705 7804: gap of unknown length
7805 8984: contig of 1180 bp in length
8985 9085: gap of unknown length
9085 10222: contig of 1138 bp in length
10223 10323: gap of unknown length
10323 11536: contig of 1204 bp in length
11537 11626: gap of unknown length
11627 12655: contig of 1029 bp in length
12656 12755: gap of unknown length
12756 13837: contig of 1082 bp in length
13838 13937: gap of unknown length
13938 15178: contig of 1241 bp in length
15179 15278: gap of unknown length
15279 17550: contig of 2272 bp in length
17551 19171: contig of 1521 bp in length
19172 19271: gap of unknown length
19272 20685: contig of 1424 bp in length
20686 20795: gap of unknown length
20796 21925: contig of 1130 bp in length
21926 22025: gap of unknown length
22026 23502: contig of 1477 bp in length
23503 23602: gap of unknown length
23603 24946: contig of 1344 bp in length
24947 25046: gap of unknown length
25047 26306: contig of 1260 bp in length
26307 26406: gap of unknown length
26407 27933: contig of 1527 bp in length
27934 28033: gap of unknown length
28034 29051: contig of 1018 bp in length
29052 29151: gap of unknown length
29152 31029: contig of 1878 bp in length
31030 31129: gap of unknown length
31130 32138: contig of 1009 bp in length
32139 32238: gap of unknown length
32240 33644: contig of 1406 bp in length
33645 33744: gap of unknown length
33745 35209: contig of 1465 bp in length
35210 35309: gap of unknown length
35310 37128: contig of 1819 bp in length
37129 37228: gap of unknown length
37229 38472: contig of 1244 bp in length
38473 38572: gap of unknown length
38573 39697: contig of 1125 bp in length
39698 39797: gap of unknown length
39799 41826: contig of 2029 bp in length
41827 41926: gap of unknown length
41927 42965: contig of 1039 bp in length
42966 43065: gap of unknown length
43066 44598: contig of 1533 bp in length
44599 44698: gap of unknown length

```

* 44699 46441: contig of 1743 bp in length
* 46542 46541: gap of unknown length
* 47797 47797: contig of 1256 bp in length
* 47898 47897: gap of unknown length
* 49123 49123: contig of 1226 bp in length
* 49223 49223: gap of unknown length
* 49224 50294: contig of 1071 bp in length
* 50295 50394: gap of unknown length
* 50395 52580: contig of 2186 bp in length
* 52581 52580: gap of unknown length
* 52681 55163: contig of 2483 bp in length
* 55164 55263: gap of unknown length
* 55264 57049: contig of 1786 bp in length
* 57050 57149: gap of unknown length
* 57150 58581: contig of 1432 bp in length
* 58582 58681: gap of unknown length
* 58682 60257: gap of 1576 bp in length
* 60258 60357: gap of unknown length
* 60358 61883: contig of 1526 bp in length
* 61884 61983: gap of unknown length
* 61984 64166: contig of 2183 bp in length
* 64167 64265: gap of unknown length
* 64267 66329: contig of 2063 bp in length
* 66330 66430: gap of unknown length
* 66430 69373: contig of 2944 bp in length
* 69374 69473: gap of unknown length
* 69474 72473: contig of 3000 bp in length
* 72474 72573: gap of unknown length
* 72574 74317: contig of 1744 bp in length
* 74318 76338: contig of 1921 bp in length
* 76339 76438: gap of unknown length
* 76439 78577: contig of 2139 bp in length
* 78578 78677: gap of unknown length
* 78678 81344: contig of 2667 bp in length
* 81345 81444: gap of unknown length
* 81445 83423: contig of 1979 bp in length
* 83424 83523: gap of unknown length
* 83524 85715: contig of 2192 bp in length
* 85716 85815: gap of unknown length
* 85816 88452: contig of 2637 bp in length
* 88453 88552: gap of unknown length
* 88553 90758: contig of 2206 bp in length
* 90759 90858: gap of unknown length
* 90859 93775: contig of 2917 bp in length
* 93776 93875: gap of unknown length
* 93876 96417: contig of 2542 bp in length

```

```

Query Match      61.2%; Score 20.8; DB 2; Length 137875;
Best Local Similarity 78.1%; Pred. No. 97;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 3 CTCCTTACTTATCGATCGGCTACTGTGAGACA 34
Db 67568 CTACTATCTTATAGACAGCTGCTGTGAGACA 67537

```

```

RESULT 5
AC109962
LOCUS
DEFINITION
AC109962 178549 bp DNA linear HTG 13-JUL-2002
Rattus norvegicus clone CH230-13SD24, *** SEQUENCING IN PROGRESS
***, 71 unordered pieces.
AC109962
VERSION
AC109962.3 GI:21738199
KEYWORDS
HTG; HTGS_PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 178549)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrooke,S.L., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Worley,K.C.
JOURNAL
TITLE
Submitted (09-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 178549)
REFERENCE
AUTHORS
Worley,K.C.
JOURNAL
TITLE
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18860226.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOLL
Center clone name: CH230-13SD24
----- Summary Statistics
Sequencing vector: Plasmid,
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 127835 bases at least Q40
Consensus quality: 133351 bases at least Q20
Consensus quality: 138582 bases at least Q20

```

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 71 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1158: contig of 1158 bp in length
 * 1159 1258: gap of unknown length
 * 1259 3027: contig of 1769 bp in length
 * 3028 3127: gap of unknown length
 * 3128 4598: contig of 1471 bp in length
 * 4599 4698: gap of unknown length
 * 4699 6401: contig of 1703 bp in length
 * 6402 6501: gap of unknown length
 * 6502 8270: contig of 1769 bp in length
 * 8271 8370: gap of unknown length
 * 8371 9753: contig of 1383 bp in length
 * 9754 9853: gap of unknown length
 * 9854 10988: contig of 1135 bp in length
 * 10989 11088: gap of unknown length
 * 11089 12826: contig of 1738 bp in length
 * 12827 12926: gap of unknown length
 * 12927 14308: contig of 1382 bp in length
 * 14309 14408: gap of unknown length
 * 14409 15711: contig of 1303 bp in length
 * 15712 15811: gap of unknown length
 * 15812 17817: contig of 2006 bp in length
 * 17818 17917: gap of unknown length
 * 17918 18963: contig of 1046 bp in length
 * 18964 19063: gap of unknown length
 * 19064 20236: contig of 1173 bp in length
 * 20237 20336: gap of unknown length
 * 20337 21839: contig of 1503 bp in length
 * 21840 21939: gap of unknown length
 * 21940 23344: contig of 1405 bp in length
 * 23345 23444: gap of unknown length
 * 23445 24472: contig of 1028 bp in length
 * 24473 24572: gap of unknown length
 * 24573 26191: contig of 1619 bp in length
 * 26192 26291: gap of unknown length
 * 26292 27406: contig of 1114 bp in length
 * 27406 27505: gap of unknown length
 * 27506 29085: contig of 1580 bp in length
 * 29086 29185: gap of unknown length
 * 29186 31257: contig of 2072 bp in length
 * 31258 31358: gap of unknown length
 * 31359 32673: contig of 1316 bp in length
 * 32674 32773: gap of unknown length
 * 32774 34608: contig of 1835 bp in length
 * 34609 34708: gap of unknown length
 * 34709 35916: contig of 1208 bp in length
 * 35917 36016: gap of unknown length
 * 36017 37116: contig of 1100 bp in length
 * 37117 37216: gap of unknown length
 * 37217 39219: contig of 1903 bp in length
 * 39220 39220: gap of unknown length
 * 39221 41424: contig of 2205 bp in length
 * 41425 41524: gap of unknown length
 * 41525 42694: contig of 1170 bp in length
 * 42695 42794: gap of unknown length
 * 42795 44400: contig of 1606 bp in length
 * 44401 44500: gap of unknown length
 * 44501 46996: contig of 2496 bp in length
 * 46997 47096: gap of unknown length
 * 47097 48371: contig of 1275 bp in length
 * 48372 48471: gap of unknown length
 * 48472 50351: contig of 1880 bp in length
 * 50352 50451: gap of unknown length
 * 50452 52100: contig of 1649 bp in length
 * 52101 52200: gap of unknown length
 * 52201 54048: contig of 1848 bp in length
 * 54049 54148: gap of unknown length
 * 54149 55721: contig of 1573 bp in length

* 55722 55821: gap of unknown length
 * 55822 57240: contig of 1419 bp in length
 * 57241 57340: gap of unknown length
 * 57341 60003: contig of 2663 bp in length
 * 60004 60103: gap of unknown length
 * 60104 62111: contig of 2008 bp in length
 * 62112 62211: gap of unknown length
 * 62212 64026: contig of 1815 bp in length
 * 64027 64126: gap of unknown length
 * 64127 66191: contig of 2065 bp in length
 * 66192 66291: gap of unknown length
 * 66292 69826: contig of 3535 bp in length
 * 69827 69926: gap of unknown length
 * 69927 71769: contig of 1843 bp in length
 * 71770 71869: gap of unknown length
 * 71870 73949: contig of 2080 bp in length
 * 73950 74049: gap of unknown length
 * 74050 76696: contig of 2647 bp in length
 * 76697 76796: gap of unknown length
 * 76797 79517: contig of 2721 bp in length
 * 79518 79617: gap of unknown length
 * 79618 81607: contig of 1990 bp in length
 * 81608 81707: gap of unknown length
 * 81708 83237: contig of 1530 bp in length
 * 83238 83337: gap of unknown length
 * 83338 85027: contig of 1690 bp in length
 * 85028 85127: gap of unknown length
 * 85128 88310: contig of 3183 bp in length
 * 88311 88410: gap of unknown length
 * 88411 92204: contig of 3794 bp in length
 * 92205 92304: gap of unknown length
 * 92305 93769: contig of 1465 bp in length
 * 93770 93869: gap of unknown length
 * 93870 96723: contig of 2853 bp in length
 * 96723 96823: gap of unknown length
 * 96823 98907: contig of 2085 bp in length
 * 98908 101249: gap of unknown length
 * 101250 101349: contig of 2242 bp in length
 * 101350 103337: contig of 1988 bp in length

Query Match 61.2%; Score 20.8; DB 2; Length 178549;
 Best Local Similarity 78.1%; Pred. No. 97;
 Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 CTCCTACTATGACGCTACGACGACA 34
 DB 77988 CTACTATCTATGACAGCGTGTGAGAGA 78019

RESULT 6
 LOCUS AC009300/c 109939 bp DNA linear PRI 30-SEP-2000
 DEFINITION Homo sapiens BAC clone RP11-52A21 from 2, complete sequence.
 ACCESSION AC009300
 VERSION AC009300.2 GI:6604538
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 109939)
 AUTHORS Sulston, J.E. and Waterston, R.
 TITLE Toward a complete human genome sequence
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
 MEDLINE 99063792
 PUBMED 9847074
 REFERENCE 2 (bases 1 to 109939)
 AUTHORS Sun, H., Wohldmann, P., Yeakum, M. and Schrader, A.
 TITLE The sequence of Homo sapiens BAC clone RP11-52A21
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 109939)
 AUTHORS Waterston, R.H.

TITLE Direct Submission
JOURNAL Submitted (13-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 4 (bases 1 to 1099339)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
Submitted (19-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 1099339)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
Submitted (21-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 6 (bases 1 to 1099339)
AUTHORS Waterston,R.
JOURNAL Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA On Dec 20, 1999 this sequence version replaced gi:5732174.

COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: saplens@wustl.wustl.edu
----- Summary Statistics8
Center project name: H_NH052A21

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,B., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-24K2, 200 bp overlap; the clone sequenced to the right is RP11-11G16. Actual start of this clone is at base position 158007 of RP11-24K2.

FEATURES
source Location/Qualifiers
1..109939
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-52A21"
/clone_id="RPCI-11"
repeat_region 157..398

/rpc_family="MIR"
repeat_region 1604..2335
/rpc_family="L2"
repeat_region 2616..2722
/rpc_family="L2"
repeat_region 3924..4253
/rpc_family="L2"
repeat_region 4508..5040
/rpc_family="L1"
repeat_region 6995..7105
/rpc_family="L2"
repeat_region 7220..7355
/rpc_family="L2"
repeat_region 7595..7769
/rpc_family="MER1_type"
repeat_region 7970..8235
/rpc_family="ALU"
repeat_region 8550..8595
/rpc_family="AT_rich"
repeat_region 9165..9511
/rpc_family="MALR"
repeat_region 10521..10829
/rpc_family="ALU"
repeat_region 11671..11977
/rpc_family="ALU"
repeat_region 13237..13374
/rpc_family="MER2_type"
repeat_region 13376..13619
/rpc_family="L1"
repeat_region 17112..17346
/rpc_family="MIR"
repeat_region 19007..19311
/rpc_family="ALU"
repeat_region 19661..19939
/rpc_family="ALU"
repeat_region 20109..20181
/rpc_family="MER1_type"
repeat_region 20650..20715
/rpc_family="MER1_type"
repeat_region 20849..20968
/rpc_family="L1"
repeat_region 20969..21344
/rpc_family="MALR"
repeat_region 21345..21969
/rpc_family="MALR"
repeat_region 21970..22338
/rpc_family="MALR"
repeat_region 22465..22712
/rpc_family="L2"
repeat_region 23563..23584
/rpc_family="AT_rich"
repeat_region 24343..24553
/rpc_family="Retroviral"
repeat_region 24971..25144
/rpc_family="Retroviral"
repeat_region 25227..25350
/rpc_family="Retroviral"
repeat_region 26618..26917
/rpc_family="ALU"
repeat_region 26949..27325
/rpc_family="MALR"
repeat_region 27758..27919
/rpc_family="L2"
repeat_region 28384..28451
/rpc_family="L2"
repeat_region 28529..29098
/rpc_family="Retroviral"
repeat_region 29473..29914
/rpc_family="L2"
repeat_region 30014..30038
/rpc_family="(TC)n"
repeat_region 30838..31133
/rpc_family="L1"

```

repeat_region      31405..31699
                    /rpt_family="Alu"
repeat_region      32303..32730
                    /rpt_family="L1"
repeat_region      32755..32829
                    /rpt_family="AT_rich"
repeat_region      32896..32962
                    /rpt_family="MER1_type"
repeat_region      33028..33311
                    /rpt_family="L1"
repeat_region      33481..34158
                    /rpt_family="L1"
repeat_region      34159..35231
                    /rpt_family="Retroviral"
repeat_region      35332..35476
                    /rpt_family="L1"
repeat_region      36141..36921
                    /rpt_family="L1"
repeat_region      37104..37147
                    /rpt_family="AT_rich"
repeat_region      38136..38717
                    /rpt_family="L1"
repeat_region      38718..39035
                    /rpt_family="Alu"
repeat_region      39036..39273
                    /rpt_family="L1"
repeat_region      39339..39650
                    /rpt_family="Alu"
repeat_region      40112..40222
                    /rpt_family="L1"
repeat_region      40231..40253
                    /rpt_family="AT_rich"
repeat_region      40440..40635

```

```

Query Match      59.4%; Score 20.2; DB 9; Length 109939;
Best Local Similarity 75.8%; Pred. No. 1.8e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Db 109638 ACTCTGTAATTAATGCTTACTGTGGACA 109606

```

```

RESULT 7
LOCUS      AC010249      111081 bp      DNA      linear      HTG 06-MAY-2000
DEFINITION Homo sapiens chromosome 5 clone CTC-406F2, WORKING DRAFT SEQUENCE.
ACCESSION  AC010249
VERSION     AC010249.5 GI:7711405
KEYWORDS    HTG; HTGS PHASE2; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 11081)
AUTHORS     Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE       Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
JOURNAL     Direct Submission
REFERENCE   2 (bases 1 to 11081)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 5
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 11081)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 6, 2000 this sequence version replaced gi:7710813.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 341206
Center clone name: CIT-HSPC_406F2

```

```

-----
Summary Statistics
Consensus quality: 104960 bases at least Q40
Consensus quality: 109090 bases at least Q30
Consensus quality: 109623 bases at least Q20
Estimated insert size: 125000; pulse field gel estimation
Estimated insert size: 109981; sum-of-contigs estimation
Quality coverage: 6.74 in Q20 bases; pulse field gel estimation
Quality coverage: 7.62 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 11274: contig of 11274 bp in length
* 11275 11374: gap of unknown length
* 11375 13418: contig of 2044 bp in length
* 13419 13518: gap of unknown length in length
* 13519 13519: contig of 22821 bp in length
* 13519 36339: gap of unknown length
* 36340 36439: gap of unknown length
* 36440 37882: contig of 1443 bp in length
* 37883 37982: gap of unknown length
* 37983 46574: contig of 8592 bp in length
* 46575 46674: gap of unknown length
* 46675 49885: contig of 3211 bp in length
* 49886 49985: gap of unknown length
* 49986 53530: contig of 3545 bp in length
* 53531 53631: gap of unknown length
* 53631 68020: contig of 14390 bp in length
* 68021 68120: gap of unknown length
* 68121 72250: contig of 4170 bp in length
* 72251 72390: gap of unknown length
* 72391 84023: contig of 11633 bp in length
* 84024 84123: gap of unknown length
* 84124 102743: contig of 18620 bp in length
* 102744 102843: gap of unknown length
* 102844 11081: contig of 8238 bp in length.
Location/Qualifiers
source
1. 111081
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-406F2"
/clone_lib="Caltech human BAC library C"
BASE COUNT      30700 a 22728 c 24378 g 32141 t 1134 others
ORIGIN

```

```

Query Match      59.4%; Score 20.2; DB 2; Length 111081;
Best Local Similarity 75.8%; Pred. No. 1.8e+02;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Db 1877 ACTCTCAATATGATCGATCGTCTGGAACA 1909

```

```

RESULT 8
LOCUS      AC011264/c      137165 bp      DNA      linear      HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-3K8, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC011264
VERSION     AC011264.4 GI:9121012
KEYWORDS    HTG; HTGS PHASE0.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 137165)
AUTHORS     Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Birren,B., Linton,L., Nuebaum,C. and Lander,B.

```


TITLE Homo sapiens, clone RP11-3K8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 137165)
AUTHORS Birren, B., Linton, L., Nuebaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barina, N., Beckery, R., Boguski, L., Bouhassira, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearfano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M., Ferris, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardys, S., Grant, G., Hago, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Kartas, A., Klein, J., Lechoczy, J., Liu, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGuire, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliou, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Center code: W1BR
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L2766
Center clone name: 3_K_8

NOTE: This record contains 161 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
651 750: contig of 100 bp in length
751 1461: contig of 711 bp in length
1462 1561: gap of 100 bp
1562 2270: contig of 709 bp in length
2271 2370: gap of 100 bp
2371 3065: contig of 695 bp in length
3066 3165: gap of 100 bp
3166 3782: contig of 617 bp in length
3783 3882: gap of 100 bp
3883 4535: contig of 653 bp in length
4536 4635: gap of 100 bp
4636 5316: contig of 681 bp in length
5317 5416: gap of 100 bp
5417 6092: contig of 676 bp in length
6093 6192: gap of 100 bp
6193 6875: contig of 683 bp in length
6876 6975: gap of 100 bp
6976 7662: contig of 687 bp in length
7663 7762: gap of 100 bp
7763 8453: contig of 691 bp in length
8454 8553: gap of 100 bp
8554 9244: contig of 691 bp in length
9245 9344: gap of 100 bp
9345 10001: contig of 657 bp in length
10002 10101: gap of 100 bp
10102 10795: contig of 694 bp in length
10796 10895: gap of 100 bp
10896 11595: contig of 700 bp in length

11596 11695: gap of 100 bp
11696 12396: contig of 701 bp in length
12397 12496: gap of 100 bp
12497 13154: contig of 658 bp in length
13155 13254: gap of 100 bp
13255 13924: contig of 670 bp in length
13925 14024: gap of 100 bp
14025 14712: contig of 688 bp in length
14713 14812: gap of 100 bp
14813 15486: contig of 674 bp in length
15487 15586: gap of 100 bp
15587 16278: contig of 692 bp in length
16279 16378: gap of 100 bp
16379 17064: contig of 686 bp in length
17065 17164: gap of 100 bp
17165 17872: contig of 708 bp in length
17873 17972: gap of 100 bp
17973 18636: contig of 664 bp in length
18637 18736: gap of 100 bp
18737 19451: contig of 715 bp in length
19452 19551: gap of 100 bp
19552 20257: contig of 706 bp in length
20258 20357: gap of 100 bp
20358 21037: contig of 680 bp in length
21038 21137: gap of 100 bp
21138 21776: contig of 639 bp in length
21777 21876: gap of 100 bp
21877 22518: contig of 642 bp in length
22519 22618: gap of 100 bp
22619 23263: contig of 645 bp in length
23264 23363: gap of 100 bp
23364 24044: contig of 681 bp in length
24045 24144: gap of 100 bp
24145 24800: contig of 656 bp in length
24801 24900: gap of 100 bp
24901 25563: contig of 663 bp in length
25564 25663: gap of 100 bp
25664 26345: contig of 682 bp in length
26346 26445: gap of 100 bp
26446 27145: contig of 700 bp in length
27146 27245: gap of 100 bp
27246 27904: contig of 659 bp in length
27905 28004: gap of 100 bp
28005 28713: contig of 709 bp in length
28714 28813: gap of 100 bp
28814 29513: contig of 700 bp in length
29514 29613: gap of 100 bp
29614 30322: contig of 709 bp in length
30323 30422: gap of 100 bp
30423 31078: contig of 656 bp in length
31079 31178: gap of 100 bp
31179 31829: contig of 651 bp in length
31830 31929: gap of 100 bp
31930 32579: contig of 650 bp in length
32580 32679: gap of 100 bp
32680 33364: contig of 685 bp in length
33365 33464: gap of 100 bp
33465 34150: contig of 686 bp in length
34151 34250: gap of 100 bp
34251 34917: contig of 667 bp in length
34918 35017: gap of 100 bp
35018 35721: contig of 704 bp in length
35722 35821: gap of 100 bp
35822 36538: contig of 717 bp in length
36539 36638: gap of 100 bp
36639 37301: contig of 663 bp in length
37302 37401: gap of 100 bp
37402 38117: contig of 716 bp in length
38118 38217: gap of 100 bp
38218 38912: contig of 695 bp in length
38913 39012: gap of 100 bp
39013 39715: contig of 703 bp in length
39716 39815: gap of 100 bp

```

* 39816 40447: contig of 632 bp in length
* 40448 40547: gap of 100 bp
* 40548 41209: contig of 662 bp in length
* 41210 41309: gap of 100 bp
* 41310 41942: contig of 633 bp in length
* 41943 42042: gap of 100 bp
* 42043 42677: contig of 635 bp in length
* 42678 42777: gap of 100 bp
* 42778 43483: contig of 706 bp in length
* 43484 43583: gap of 100 bp
* 43584 44276: contig of 693 bp in length
* 44277 44376: gap of 100 bp
* 44377 45058: contig of 682 bp in length
* 45059 45158: gap of 100 bp
* 45159 45820: contig of 662 bp in length
* 45821 45920: gap of 100 bp
* 45921 46631: contig of 711 bp in length
* 46632 46731: gap of 100 bp
* 46732 47426: contig of 695 bp in length
* 47427 47526: gap of 100 bp
* 47527 48239: contig of 713 bp in length
* 48240 48339: gap of 100 bp
* 48340 48988: contig of 649 bp in length
* 48989 49088: gap of 100 bp
* 49089 49770: contig of 682 bp in length
* 49771 49870: gap of 100 bp
* 49871 50553: contig of 683 bp in length
* 50554 50653: gap of 100 bp
* 50654 51336: contig of 683 bp in length
* 51337 51436: gap of 100 bp
* 51437 52109: contig of 673 bp in length
* 52110 52209: gap of 100 bp
* 52210 52900: contig of 691 bp in length
* 52901 53000: gap of 100 bp
* 53001 53714: contig of 714 bp in length
* 53715 53814: gap of 100 bp
* 53815 54443: contig of 629 bp in length
* 54444 54543: gap of 100 bp
* 54544 55242: contig of 699 bp in length
* 55243 55342: gap of 100 bp
* 55343 56046: contig of 704 bp in length
* 56047 56146: gap of 100 bp
* 56147 56834: contig of 688 bp in length

```

Query Match 59.4%; Score 20.2; DB 2; Length 137165;
 Best Local Similarity 75.8%; Pred. No. 1.8e+02;
 Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 ACTCCTACTTATCGATCGGTACTGTGAGACA 34
 Db 7430 ACTGCTCCACATATCGATCGGTCTTCTAGAAACA 7398

RESULT 9
 AC026799/c 162918 bp DNA linear PRI 11-AUG-2001
 LOCUS Homo sapiens chromosome 5 clone CTD-2366H10, complete sequence.
 DEFINITION AC026799
 ACCESSION AC026799.4 GI:15148112
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 162918)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Aug 11, 2001 this sequence version replaced gi:13677048:
 draft sequence produced by DOE Joint Genome Institute
 www.jgi.doe.gov

FEATURES
 source
 1. 162918
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CTD-2366H10"
 location/Qualifiers
 Estimated total Number of Errors is 0.2.

BASE COUNT 47872 a 35261 c 34959 g 44826 t
 ORIGIN

Query Match 59.4%; Score 20.2; DB 9; Length 162918;
 Best Local Similarity 75.8%; Pred. No. 1.8e+02;
 Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 ACTCCTACTTATCGATCGGTACTGTGAGACA 34
 Db 108854 ACTGCTCCACATATCGATCGGTCTTCTAGAAACA 108822

RESULT 10
 AC121165/c 176172 bp DNA linear HTG 18-JUL-2002
 LOCUS Rattus norvegicus clone CH230-129C5, *** SEQUENCING IN PROGRESS
 DEFINITION AC121165
 ACCESSION AC121165.2 GI:21747197
 VERSION
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Norway rat.
 ORGANISM

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 176172)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alstbrooks,S.L., Anazatunge,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbieri,J., Benton,D., Bimge,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhaq,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,C., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falle,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Prantz,P.,
 Gabisi,A., Geo,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Goriell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Haylak,P., Hayes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Hulyk,S., Hune,T., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolyvet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
 Lozador,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mashiney,E., McLeod,M.P., Meador,K., Morgan,M., Morris,S.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,N., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenko,S., Oguh,M., Okwona,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,D., Perez,L.,

Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,
Rivera, M., Rojase, A., Rojasek, I., Rolfe, M., Ruiz, S., Savery, G.,
Scheer, S., Scott, G., Shen, H., Shoochhari, N., Sisson, I.,
Sodegren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Taber, P., Tameria, A., Tameria, K., Tang, H.,
Tansy, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Uemami, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176172)
AUTHORS Worley, K.C.
JOURNAL Direct Submission
Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 176172)
Worley, K.C.
Direct Submission
Submitted (18-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 14, 2002 this sequence version replaced gi:20806189.

COMMENT
JOURNAL
AUTHORS
TITLE
JOURNAL
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: GVRC
Center clone name: CH230-129C5
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 115261 bases at least Q40
Consensus quality: 121020 bases at least Q30
Consensus quality: 124933 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1118: contig of 1118 bp in length
1 1119 1218: gap of unknown length
1 1219 2651: contig of 1433 bp in length
1 2652 2751: gap of unknown length
1 2752 4118: contig of 1367 bp in length
1 4119 4219: gap of unknown length
1 4219 5515: contig of 1297 bp in length
1 5516 5615: gap of unknown length
1 5616 6879: contig of 1264 bp in length
1 6880 6979: gap of unknown length
1 6980 8084: contig of 1105 bp in length
1 8085 8184: gap of unknown length
1 8185 9269: contig of 1085 bp in length
1 9270 9369: gap of unknown length
1 9370 10868: contig of 1499 bp in length
1 10869 12116: contig of 1148 bp in length
1 12117 12216: gap of unknown length
1 12217 13432: contig of 1216 bp in length
1 13433 13533: gap of unknown length
1 13533 14843: contig of 1311 bp in length

14944 14943: gap of unknown length
14944 16236: contig of 1293 bp in length
16237 16336: gap of unknown length
16337 17557: contig of 1221 bp in length
17558 17657: gap of unknown length
17658 18680: contig of 1023 bp in length
18681 18780: gap of unknown length
18781 19823: contig of 1043 bp in length
19824 19923: gap of unknown length
19924 21040: contig of 1117 bp in length
21041 21140: gap of unknown length
21141 22332: contig of 1192 bp in length
22333 22432: gap of unknown length
22433 23439: contig of 1007 bp in length
23440 24639: gap of unknown length
24640 24739: contig of 1100 bp in length
24740 26190: gap of unknown length
26191 26290: contig of 1451 bp in length
26291 27930: gap of unknown length
27931 28030: contig of 1640 bp in length
28031 29345: gap of unknown length
29346 29445: contig of 1315 bp in length
29446 30567: gap of unknown length
30568 32667: contig of 1122 bp in length
32668 33268: gap of unknown length
33269 33885: contig of 1601 bp in length
33886 33985: gap of unknown length
33986 35460: contig of 1517 bp in length
35461 35560: gap of unknown length
35561 37289: contig of 1475 bp in length
37290 37389: gap of unknown length
37390 39446: contig of 1729 bp in length
39447 39546: gap of unknown length
39547 41715: contig of 2057 bp in length
41716 43126: gap of unknown length
43127 44501: contig of 2169 bp in length
44502 44601: gap of unknown length
44602 46257: contig of 1275 bp in length
46258 46357: gap of unknown length
46358 48457: contig of 1656 bp in length
48458 48557: gap of unknown length
48559 50541: contig of 2100 bp in length
50542 50641: gap of unknown length
50642 52901: contig of 1984 bp in length
52902 53001: gap of unknown length
53002 55713: contig of 2260 bp in length
55714 57629: gap of unknown length
57630 57729: contig of 2712 bp in length
57730 59173: gap of unknown length
59174 59272: contig of 1816 bp in length
59273 61015: gap of unknown length
61016 63508: contig of 1443 bp in length
63509 63608: gap of unknown length
63609 65779: contig of 1742 bp in length
65780 68016: gap of unknown length
68017 68117: contig of 2394 bp in length
68118 71085: gap of unknown length
71086 71185: contig of 2969 bp in length
71186 73899: gap of unknown length
73900 76690: contig of 2714 bp in length
76691 76790: gap of unknown length
76791 79764: contig of 2691 bp in length
79765 79864: gap of unknown length
79865 82315: contig of 2974 bp in length
82316 82414: gap of unknown length

```

*      82415      85631: contig of 3217 bp in length
*      85632      85731: gap of unknown length
*      85732      87892: contig of 2161 bp in length
*      87893      87992: gap of unknown length
*      87993      89863: contig of 1871 bp in length
*      89864      89963: gap of unknown length
*      89964      93583: contig of 3620 bp in length
*      93584      93684: gap of unknown length
*      93685      96249: contig of 2566 bp in length
*      96250      96349: gap of unknown length
*      96350      98756: contig of 2407 bp in length
*      98757      98856: gap of unknown length
*      98857      99883: contig of 1027 bp in length

Query Match
Best Local Similarity 75.8%; Pred. No. 1.8e+02; Length 176172;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY      1      GACTCCTTACTTATCGATCGTACTGTAGAC 33
Db 122052 GTCCTCTTACACGACGACTGTACTGTAGAC 122020

RESULT 11
AC022438/c 185791 bp DNA linear HTG 04-OCT-2001
LOCUS Homo sapiens chromosome 5 clone RP11-137P5, WORKING DRAFT SEQUENCE,
DEFINITION 9 ordered pieces.
ACCESSION AC022438
VERSION AC022438.4 GI:14579711
KEYWORDS HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 185791)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 185791)
TITLE DOE Joint Genome Institute.
AUTHORS Direct Submission
JOURNAL Submitted (03-FEB-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 1, 2001 this sequence version replaced gi:7711740.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 459338
Center clone name: RP11-137P5
-----
Summary Statistics
Consensus quality: 181059 bases at least Q40
Consensus quality: 183500 bases at least Q30
Consensus quality: 184433 bases at least Q20
Estimated insert size: 198400; agarose-fp estimation
Estimated insert size: 184991; sum-of-contigs estimation
Quality coverage: 7.42 in Q20 bases; agarose-fp estimation
Quality coverage: 7.96 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 71478: contig of 71478 bp in length
* 71479 71578: gap of unknown length

```

```

*      71579      80602: contig of 9024 bp in length
*      80603      80702: gap of unknown length
*      80703      93617: contig of 12915 bp in length
*      93618      93717: gap of unknown length
*      93718      116162: contig of 22445 bp in length
*      116163      116262: gap of unknown length
*      116263      118116: contig of 1854 bp in length
*      118117      118216: gap of unknown length
*      118217      137813: contig of 19597 bp in length
*      137814      137913: gap of unknown length
*      137914      139655: contig of 1742 bp in length
*      139656      139755: gap of unknown length
*      139756      177952: contig of 38197 bp in length
*      177953      178052: gap of unknown length
*      178053      185791: contig of 7739 bp in length.

FEATURES
Source
1. 185791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-137P5"
/clone_11b="RP11 human BAC library 11"
BASE COUNT 53841 a 39773 c 38649 g 52728 t 800 others
ORIGIN
Query Match
Best Local Similarity 75.8%; Pred. No. 1.8e+02; Length 185791;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY      2      ACTCCTTACTTATCGATCGTACTGTAGACA 34
Db 66938 ACTCCTCACATATCGATCGTCTCTAGAAACA 66906

RESULT 12
AC130200/c 80229 bp DNA linear HTG 19-AUG-2002
LOCUS Medicago truncatula clone mchl-64n13, WORKING DRAFT SEQUENCE, 28
DEFINITION unordered pieces.
ACCESSION AC130200
VERSION AC130200.5 GI:22297341
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE barcel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 80229)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Summer,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Medicago truncatula BAC Clone mchl-64n13
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 80229)
TITLE Medicago truncatula BAC Clone mchl-64n13
AUTHORS Direct Submission
JOURNAL Submitted (08-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 80229)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Summer,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Aug 19, 2002 this sequence version replaced gi:22218474.
-----Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 28 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1999: contig of 1999 bp in length
2000 2099: gap of unknown length
4226: contig of 2127 bp in length
4227 4326: gap of unknown length
4327 6630: contig of 2304 bp in length
6631 6730: gap of unknown length
6731 9394: contig of 2664 bp in length
9395 9494: gap of unknown length
12005 12006: contig of 2511 bp in length
12105 12106: gap of unknown length
14839 14840: contig of 2734 bp in length
14939 14940: gap of unknown length
17314 17315: contig of 2375 bp in length
17414 17415: gap of unknown length
19640 19641: contig of 2226 bp in length
19740 19741: gap of unknown length
21813 21814: contig of 2073 bp in length
25010 25011: contig of 3097 bp in length
25110 25111: gap of unknown length
28032 28033: contig of 2922 bp in length
28132 28133: gap of unknown length
30281 30282: contig of 2149 bp in length
30381 30382: gap of unknown length
33040 33041: contig of 2659 bp in length
33140 33141: gap of unknown length
35458 35459: contig of 2318 bp in length
35558 35559: gap of unknown length
37727 37728: contig of 2169 bp in length
37827 37828: gap of unknown length
39889 39890: contig of 2062 bp in length
39989 39990: gap of unknown length
42308 42309: contig of 2319 bp in length
42408 42409: gap of unknown length
44708 44709: contig of 2300 bp in length
44808 44809: gap of unknown length
47075 47076: contig of 2267 bp in length
47175 47176: gap of unknown length
50293 50294: contig of 3118 bp in length
50393 50394: gap of unknown length
52798 52799: contig of 2405 bp in length
52898 52899: gap of unknown length
57000 57001: contig of 4102 bp in length
57100 57101: gap of unknown length
60039 60040: contig of 2939 bp in length
60139 60140: gap of unknown length
63242 63243: contig of 3103 bp in length
63342 63343: gap of unknown length
66407 66408: contig of 3065 bp in length
66507 66508: gap of unknown length
72073 72074: contig of 5566 bp in length
72173 72174: gap of unknown length
77199 77200: contig of 5026 bp in length
77299 77300: gap of unknown length
80229 80230: contig of 2930 bp in length.
Location/Qualifiers
1. 80229
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_1ib="Medicago truncatula BAC library H1"

```

BASE COUNT 25729 a 13595 c 12936 g 25254 t 2715 others

Query Match 58.8%; Score 20; DB 2; Length 80229;

Best Local Similarity 82.1%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 ACTCCCTACTTATCGATCGTACTG 29
Db 43490 ACTCTCACTTTGATCGTACTG 43463

```

RESULT 13
AC121899/c 159746 bp DNA linear HTG 21-MAY-2002
LOCUS AC121899
DEFINITION Mus musculus chromosome UNK clone RP24-166J12, WORKING DRAFT
SEQUENCE 4 unordered pieces.
ACCESSION AC121899
VERSION AC121899.1 GI:21040020
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 159746)
AUTHORS McPherson,J.D. and Waterston,R.H.
JOURNAL The sequence of Mus musculus clone
REFERENCE 2 (bases 1 to 159746)
AUTHORS McPherson,J.D. and Waterston,R.H.
JOURNAL Unpublished
TITLE Direct Submission
COMMENT Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

```

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WTGSC
Web site: http://genome.wustl.edu/gac/index.shtml
Contact: submissions@wustl.wustl.edu
Project information
Center project name: M B8016J12
----- Summary Statistics -----
Sequencing vector: MJ3, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 15759 bases at least Q40
Consensus quality: 15836 bases at least Q30
Insert size: 159446; agarose-fp
Quality coverage: 16.21 in Q20 bases; agarose-fp
Quality coverage: 12.58 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1 8021: contig of 8021 bp in length
2 8022 8121: gap of unknown length
3 8122 26511: contig of 18390 bp in length
4 26512 26513: gap of unknown length
5 26514 76565: contig of 49954 bp in length
6 76566 76567: gap of unknown length
7 76568 159746: contig of 83081 bp in length.
Location/Qualifiers
1. 159746
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-166J12"

```

FEATURES

source

```

misc_feature      1..8021
                  /note="assembly_name:Contig12"
misc_feature      8122..26511
                  /note="assembly_name:Contig13"
misc_feature      26612..76565
                  /note="assembly_name:Contig14"
misc_feature      76666..159746
                  /note="assembly_name:Contig15"
BASE COUNT      51030 a 29319 c 30906 g 48183 t 308 others
ORIGIN
Query Match      58.84; Score 20; DB 2; Length 159746;
Best Local Similarity 82.14; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Db 81057 TCTACTTATCTCTCTGTCTTGAGACA 81030

```

```

RESULT 14
LOCUS      AC107106      170083 bp      DNA      linear      HTG 17-JUL-2002
DEFINITION      Rattus norvegicus clone CH230-155M2. *** SEQUENCING IN PROGRESS
ACCESSION      AC107106.3 GI:21743917
VERSION        HTG; HTGS PHASE1.
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Wuzy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Albrooke,S.L., Amaralung,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbata,J., Benton,J., Bimege,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeJany,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falle,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabibi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homesi,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karibson,B., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Loulesged,H.,
Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Mashwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Monabadi,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okumotu,G.,
Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Prime,E., Pu,L.L., Quilez,M., Ran,Y.,
Rivas,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Syatek,A., Tabor,P., Tameria,A., Tamerias,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Tumati,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wlaczky,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Worley,K.C.
JOURNAL
REFERENCE
AUTHORS
Worley,K.C.
JOURNAL
COMMENT
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 170083)
Worley,K.C.
Direct Submission
Submitted (16-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 170083)
Worley,K.C.
Direct Submission
Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:18701531.
-----
Center: Baylor College of Medicine
Genome Center
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project information
Center project name: GNIF
Center clone name: CH230-155M2
-----
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990129
Consensus quality: 111369 bases at least Q40
Consensus quality: 116447 bases at least Q30
Consensus quality: 121122 bases at least Q20
-----

```

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1039: contig of 1039 bp in length
1040 1139: gap of unknown length
1140 2259: contig of 1120 bp in length
2260 2359: gap of unknown length
2360 3500: contig of 1141 bp in length
3501 3600: gap of unknown length
3601 4883: contig of 1283 bp in length
4884 4983: gap of unknown length
4984 6335: contig of 1352 bp in length
6336 6435: gap of unknown length
6436 7741: contig of 1306 bp in length
7742 7841: gap of unknown length
7842 9151: contig of 1310 bp in length
9152 9251: gap of unknown length
9252 10753: contig of 1502 bp in length
10754 10853: gap of unknown length
10854 12543: contig of 1690 bp in length
12544 12643: gap of unknown length
12644 13865: contig of 1222 bp in length
13866 13965: gap of unknown length
13966 15202: contig of 1231 bp in length
15203 15302: gap of unknown length
15303 16317: contig of 1015 bp in length
16318 16417: gap of unknown length
16418 17872: contig of 1455 bp in length
17873 17972: gap of unknown length
17973 19199: contig of 1227 bp in length
19200 19299: gap of unknown length
20563: contig of 1264 bp in length
20663: gap of unknown length
22126: contig of 1463 bp in length

```

```

22127 22226: gap of unknown length
22227 23739: contig of 1513 bp in length
22740 23839: gap of unknown length
23840 25334: contig of 1495 bp in length
25335 25434: gap of unknown length
25435 27296: contig of 1862 bp in length
27297 27396: gap of unknown length
27397 28964: contig of 1567 bp in length
28964 29062: gap of unknown length
30623 30723: contig of 1559 bp in length
30723 31939: gap of unknown length
31939 32039: contig of 1217 bp in length
32039 34250: gap of unknown length
34250 34350: contig of 2211 bp in length
34351 36580: gap of unknown length
36581 36680: contig of 2230 bp in length
36681 38029: gap of unknown length
38029 38129: contig of 1349 bp in length
38130 40235: gap of unknown length
40235 40336: contig of 2106 bp in length
40336 42225: gap of unknown length
42225 42325: contig of 1890 bp in length
42326 44117: gap of unknown length
44117 44217: contig of 1792 bp in length
44218 44217: gap of unknown length
44218 45978: contig of 1761 bp in length
45979 46078: gap of unknown length
46079 47659: contig of 1581 bp in length
47660 47759: gap of unknown length
47760 49260: contig of 1501 bp in length
49261 49360: gap of unknown length
49361 51125: contig of 1765 bp in length
51126 51225: gap of unknown length
51226 54075: contig of 2851 bp in length
54077 54175: gap of unknown length
54177 55760: contig of 1584 bp in length
55761 55860: gap of unknown length
55861 58528: contig of 2668 bp in length
58529 58628: gap of unknown length
58629 61885: contig of 3257 bp in length
61886 61985: gap of unknown length
61986 65195: contig of 3210 bp in length
65196 65295: gap of unknown length
65296 67735: contig of 2340 bp in length
67736 70102: gap of unknown length
70103 70202: gap of unknown length
70203 73728: contig of 3526 bp in length
73729 73828: gap of unknown length
73829 76882: contig of 3054 bp in length
76883 80139: gap of unknown length
80140 80239: gap of unknown length
80240 82873: contig of 2634 bp in length
82874 82973: gap of unknown length
82974 85025: contig of 2052 bp in length
85026 85125: gap of unknown length
85126 89495: contig of 4370 bp in length
89496 89595: gap of unknown length
89596 92336: contig of 2741 bp in length
92337 92436: gap of unknown length
92437 95301: contig of 2865 bp in length
95302 95401: gap of unknown length
95402 100512: contig of 5111 bp in length
100513 100612: gap of unknown length
100613 103557: contig of 2945 bp in length
103558 103657: gap of unknown length
103659 107805: contig of 4148 bp in length
107806 107905: gap of unknown length
107906 110962: contig of 3057 bp in length
110963 111062: gap of unknown length
111063 114462: contig of 3400 bp in length
114463 114562: gap of unknown length

```

```

* 114563 117995: contig of 3433 bp in length
* 117996 118095: gap of unknown length
* 118096 121626: contig of 3531 bp in length

Query Match 58.8%; Score 20; DB 2; Length 170083;
Best Local Similarity 82.1%; Pred. No. 2.3e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 4 TCCCTACTATCATCGACTGCTGAG 31
Db 60176 TCCATCTACTATCATGATGACTGGAAG 60203

```

```

RESULT 15
AC129318/c
LOCUS
DEFINITION
Mus musculus chromosome UNK clone RP23-464B17, WORKING DRAFT
AC129318
AC129318.1 GI:22000919
VERSION
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
house mouse.
SOURCE
Mus musculus
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 227789)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
2 (bases 1 to 227789)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (28-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watscn.wustl.edu
Center project name: M_BA0464B17

----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 222687 bases at least Q40
Consensus quality: 223874 bases at least Q30
Consensus quality: 224621 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 226459; sum-of-contigs
Quality coverage: 17.88 in Q20 bases; agarose-fp
Quality coverage: 11.92 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1364: contig of 1364 bp in length
* 1365 1464: gap of unknown length
* 1465 3160: contig of 1696 bp in length
* 3161 3260: gap of unknown length
* 3261 5330: contig of 2060 bp in length
* 5331 5420: gap of unknown length
* 5421 6609: contig of 1189 bp in length
* 6610 6709: gap of unknown length

```

```

*      6710      24323: contig of 17614 bp in length
*      24324      24423: gap of unknown length
*      24424      31815: contig of 7393 bp in length
*      31817      31915: gap of unknown length
*      31917      42488: contig of 10572 bp in length
*      42489      42588: gap of unknown length
*      42589      42718: contig of 130 bp in length
*      42719      42818: gap of unknown length
*      42819      73139: contig of 30321 bp in length
*      73140      73239: gap of unknown length
*      73240      97703: contig of 24464 bp in length
*      97704      97803: gap of unknown length
*      97804      137021: contig of 39218 bp in length
*      137022      180567: gap of unknown length
*      137122      180567: contig of 43446 bp in length
*      180568      227789: gap of unknown length
*      180668      227789: contig of 47122 bp in length.

```

FEATURES

Location/Qualifiers

1..227789

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="UNK"

/clone="RP23-464B17"

1..1364

/note="assembly_name:Contig49"

1465..3160

/note="assembly_name:Contig51"

3261..5320

/note="assembly_name:Contig52"

5421..6609

/note="assembly_name:Contig53"

6710..24323

/note="assembly_name:Contig54"

24424..31816

/note="assembly_name:Contig55"

31917..42488

/note="assembly_name:Contig56"

42589..42718

/note="assembly_name:Contig57"

42819..73139

/note="assembly_name:Contig58"

73240..97703

/note="assembly_name:Contig59"

97804..137021

/note="assembly_name:Contig60"

137122..180567

/note="assembly_name:Contig61"

180668..227789

/note="assembly_name:Contig62"

BASE COUNT 70688 a 40836 c 41653 g 73395 t 1217 others

ORIGIN

Query Match 58.8%; Score 20; DB 2; Length 227789;

Best Local Similarity 82.1%; Pred. No. 2.2e+02;

Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 TCTACTTATCGATCGTACTGAGACA 34
 |||||
 Db 11101 TCTACTTATCTCTGTTCTTGAGACA 11074

Search completed: May 11, 2003, 06:08:20;
 Job time: 558.11 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 02:48:15 ; Search time 1269.94 Seconds
(without alignments)
1007.484 Million cell updates/sec

Title: US-09-963-803-14

Perfect score: 79
Sequence: 1 catgctgcagactatgact.....aaggatgacacatgccact 79

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vtc:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	54.4	224	17	BH746858
2	39	49.4	112	17	BH751161
3	39	49.4	142	17	BH749349
4	39	49.4	153	17	BH619283
5	39	49.4	153	17	BH747013
6	39	49.4	158	17	BH748291

Result	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
7	39	49.4	165	17	BH747357							
8	39	49.4	175	17	BH746616							
9	39	49.4	177	17	BH802465							
10	39	49.4	190	17	BH752801							
11	39	49.4	193	17	BH748289							
12	39	49.4	214	17	BH747829							
13	39	49.4	215	17	BH753813							
14	39	49.4	219	17	BH747744							
15	39	49.4	220	17	BH747438							
16	39	49.4	221	17	BH802415							
17	39	49.4	220	17	BH746474							
18	39	49.4	230	17	BH799173							
19	39	49.4	237	17	BH802463							
20	39	49.4	244	17	BH746375							
21	39	49.4	248	17	BH254798							
22	39	49.4	251	17	BH750171							
23	39	49.4	254	17	BH748500							
24	39	49.4	256	17	BH748499							
25	39	49.4	258	17	BH802470							
26	39	49.4	261	17	BH802493							
27	39	49.4	268	17	BH611919							
28	39	49.4	269	17	BH211546							
29	39	49.4	269	17	BH802489							
30	39	49.4	271	17	BH799178							
31	39	49.4	272	17	BH802428							
32	39	49.4	274	17	BH802495							
33	39	49.4	279	17	BH748475							
34	39	49.4	281	17	BH750170							
35	39	49.4	281	17	BH802443							
36	39	49.4	284	17	BH213307							
37	39	49.4	284	17	BH746665							
38	39	49.4	286	17	BH747007							
39	39	49.4	293	17	BH610310							
40	39	49.4	296	17	BH213532							
41	39	49.4	335	17	BH213230							
42	39	49.4	344	17	BH747662							
43	39	49.4	350	17	BH747675							
44	39	49.4	355	17	BH747371							
45	39	49.4	360	17	BH747123							

ALIGNMENTS

RESULT 1
LOCUS BH746858 224 bp. DNA linear GSS 27-FEB-2002
DEFINITION SALK_003694.51.40.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_003694.51.40.x. DNA
sequence.
BH746858
VERSION BH746858.1 GI:18959973
KEYWORDS GSS.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 224)
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeke, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.
A Sequence-indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of

DEFINITION SALK_040762 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_040762, DNA sequence.

ACCESSION BH619283

VERSION BH619283.1 GI:18429799

KEYWORDS GSS

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 153)

AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

FEATURES
source
Class: TDNA tagged.
Location/Qualifiers
1..153
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_040762"
/note="11b="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT 36 a 34 c 36 g 47 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 153;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72
|||||
Db 110 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72

RESULT 5
BH747013 153 bp DNA linear GSS 27-FEB-2002

LOCUS SALK_008070.43.05.x Arabidopsis thaliana TDNA insertion lines

DEFINITION Arabidopsis thaliana genomic clone SALK_008070.43.05.x, DNA sequence.

ACCESSION BH747013

VERSION BH747013.1 GI:18960128

KEYWORDS GSS

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 153)

AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

FEATURES
source
Class: TDNA tagged.
Location/Qualifiers
1..153
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_008070.43.05.x"
/note="11b="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT 47 a 37 c 33 g 36 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 153;
Best Local Similarity 100.0%; Pred. No. 0.0035;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72
|||||
Db 44 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 82

RESULT 6
BH748291 158 bp DNA linear GSS 27-FEB-2002

LOCUS SALK_045100.51.10.x Arabidopsis thaliana TDNA insertion lines

DEFINITION Arabidopsis thaliana genomic clone SALK_045100.51.10.x, DNA sequence.

ACCESSION BH748291

VERSION BH748291.1 GI:18961648

KEYWORDS GSS

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 158)

AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

FEATURES
source
Class: TDNA tagged.
Location/Qualifiers
1..158
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_045100.51.10.x"
/clone_11b="Arabidopsis thaliana TDNA insertion lines"

/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT

37 a 35 c 37 g 49 t

Query Match

49.4%; Score 39; DB 17; Length 158;

Best Local Similarity 100.0%; Pred. No. 0.0036; Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72
DB 110 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

RESULT 7

BH747357 165 bp DNA linear GSS 27-FEB-2002

LOCUS SALK_016522.55.50.x Arabidopsis thaliana TDNA insertion lines

DEFINITION Arabidopsis thaliana genomic clone SALK_016522.55.50.x, DNA

sequence.

ACCESSION BH747357

VERSION BH747357

KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 165)

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

Unpublished (2001)

CONTACT: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA.

Class: TDNA tagged.

Location/Qualifiers

1. 165 /organism="Arabidopsis thaliana"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_016522.55.50.x"

/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT 45 a 43 c 39 g 38 t

ORIGIN

Query Match

49.4%; Score 39; DB 17; Length 165;

Best Local Similarity 100.0%; Pred. No. 0.0036; Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

DB 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

DB 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

RESULT 8

BH746616

LOCUS

DEFINITION

Arabidopsis thaliana genomic clone SALK_045268.54.50.x, DNA

sequence.

ACCESSION BH746616

VERSION BH746616

KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 175)

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

Unpublished (2001)

CONTACT: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA.

Class: TDNA tagged.

Location/Qualifiers

1. 175 /organism="Arabidopsis thaliana"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_045268.54.50.x"

/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/cdna_protocols.html"

BASE COUNT 54 a 43 c 40 g 38 t

ORIGIN

Query Match

49.4%; Score 39; DB 17; Length 175;

Best Local Similarity 100.0%; Pred. No. 0.0038; Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

DB 66 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 104

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

DB 66 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 104

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

DB 66 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 104

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

DB 66 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 104

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

DB 66 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 104

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

DB 66 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 104

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

DB 66 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 104

Y 34 GATTGATGTGATATCTCCACTGACGTAAGGATGACGCA 72

Maize genomic sequences found using engineered Rescued transposon

JOURNAL
COMMENT

Unpublished (2001)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Very probable ligation site found so sequence was trimmed.
Post-ligation sequence submitted separately.
Plate: 1008026 row: 10
Class: transposon-tagged.

FEATURES
Source

Location/Qualifiers
1. 177
/organism="Zea mays"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/clone_1id="1008 - RescueMu Grid 1"
/issue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: leaf; Vector: RescueMu (engineered from pBluescript backbone); Site 1: BamHI, Site 2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site www.zmdb.iastate.edu and follow the links for 'RescueMu.' Grid 1 was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT
39 a 38 c 47 g 53 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 177;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATACGCA 72
|||||
Db 118 GATTGATGTGATATCTCCACTGACGTAAAGGATACGCA 80
|||||

RESULT 10
LOCUS BH752801 190 bp DNA linear GSS 27-FEB-2002
DEFINITION SALK_019366.54.25.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_019366.54.25.x, DNA
sequence.

ACCESSION BH752801
VERSION BH752801.1 GI:18972023
KEYWORDS GSS.
SOURCE
ORGANISM

Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
1 (bases 1 to 190)
, C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
, Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the

TITLE
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

JOURNAL
COMMENT

Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.

FEATURES
Source

Location/Qualifiers
1. 190
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone_1id="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT
52 a 52 c 41 g 45 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 190;
Best Local Similarity 100.0%; Pred. No. 0.0039;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATACGCA 72
|||||
Db 81 GATTGATGTGATATCTCCACTGACGTAAAGGATACGCA 119
|||||

RESULT 11
LOCUS BH748289 193 bp DNA linear GSS 27-FEB-2002
DEFINITION SALK_045097.54.25.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_045097.54.25.x, DNA
sequence.

ACCESSION BH748289
VERSION BH748289.1 GI:18961643
KEYWORDS GSS.
SOURCE
ORGANISM

Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
1 (bases 1 to 193)
, C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
, Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the

TITLE
Arabidopsis Genome
Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.

FEATURES
Source

Location/Qualifiers
1. 193
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone_1id="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT
61 a 48 c 43 g 41 t

Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 193;
 Best Local Similarity 100.0%; Pred. No. 0.004;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
 |||||
 DB 84 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 122

RESULT 12

BH747829

LOCUS BH747829 214 bp DNA linear GSS 27-FEB-2002
 DEFINITION SALK_035354.55.75.x Arabidopsis thaliana TDNA insertion lines
 Arabidopsis thaliana genomic clone SALK_035354.55.75.x, DNA
 sequence.

ACCESSION BH747829

VERSION BH747829

KEYWORDS GI:18960944

SOURCE GSS.

ORGANISM Arabidopsis thaliana

thale cress.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

1 (bases 1 to 214)
 Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadriab
 ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Predits,L., Shinn,P.
 , Zimmerman,J. and Ecker,J.R.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome

JOURNAL Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu

This is single pass sequence recovered from the left border of
 TDNA.

Class: TDNA tagged.
 Location/Qualifiers

FEATURES

source

1..214
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_035354.55.75.x"
 /note="lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"
 BASE COUNT 58 a 56 c 52 g 48 t

Query Match 49.4%; Score 39; DB 17; Length 214;
 Best Local Similarity 100.0%; Pred. No. 0.0042;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
 |||||
 DB 104 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 142

RESULT 13

BH753813

LOCUS BH753813 215 bp DNA linear GSS 27-FEB-2002
 DEFINITION SALK_029677.55.00.x Arabidopsis thaliana TDNA insertion lines
 Arabidopsis thaliana genomic clone SALK_029677.55.00.x, DNA
 sequence.

ACCESSION BH753813
 VERSION BH753813.1 GI:18974308
 KEYWORDS GSS.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

1 (bases 1 to 215)
 Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadriab
 ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Predits,L., Shinn,P.
 , Zimmerman,J. and Ecker,J.R.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome

JOURNAL

COMMENT

Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu

This is single pass sequence recovered from the left border of
 TDNA.

Class: TDNA tagged.
 Location/Qualifiers

FEATURES

source

1..215
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_029677.55.00.x"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"
 BASE COUNT 63 a 56 c 50 g 46 t

Query Match 49.4%; Score 39; DB 17; Length 215;
 Best Local Similarity 100.0%; Pred. No. 0.0042;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 72
 |||||
 DB 105 GATTGATGTGATATCTCCACTGACGTAAAGGATGACGCA 143

RESULT 14

BH747744

LOCUS BH747744 219 bp DNA linear GSS 27-FEB-2002
 DEFINITION SALK_033710.53.75.x Arabidopsis thaliana TDNA insertion lines
 Arabidopsis thaliana genomic clone SALK_033710.53.75.x, DNA
 sequence.

ACCESSION BH747744

VERSION BH747744.1

KEYWORDS GI:18960859

SOURCE GSS.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

1 (bases 1 to 219)
 Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadriab
 ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Predits,L., Shinn,P.
 , Zimmerman,J. and Ecker,J.R.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome

JOURNAL

COMMENT

Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.
Location/Qualifiers
1. 219

FEATURES
source
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_033710.53.75.x"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT
66 a 57 c 50 g 46 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 219;
Best Local Similarity 100.0%; Pred. No. 0.0043;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72
|||||
Db 109 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 147

RESULT 15

LOCUS

BH747438 220 bp DNA linear GSS 27-FEB-2002
SALK_017469.51.60.x Arabidopsis thaliana TDNA insertion lines

DEFINITION

Arabidopsis thaliana genomic clone SALK_017469.51.60.x, DNA

ACCESSION

BH747438

VERSION

BH747438.1

KEYWORDS

GSS.

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

REFERENCE

Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosoids II; Brassicales; Brassicaceae; Arabidopsis.

AUTHORS

1 (bases 1 to 220)
Alonso, J.M., Leisner, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrihab,
C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.,
Zimmerman, J. and Ecker, J.R.

TITLE

A Sequence-Indexed Library of Insertion Mutations in the

JOURNAL

Arabidopsis Genome
Unpublished (2001)

COMMENT

Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)

The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

TDNA.

FEATURES

Class: TDNA tagged.
Location/Qualifiers
1. 220

source

/organism="Arabidopsis thaliana"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_017469.51.60.x"

/note="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT
67 a 55 c 48 g 50 t

ORIGIN

Query Match 49.4%; Score 39; DB 17; Length 220;
Best Local Similarity 100.0%; Pred. No. 0.0043;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 72
|||||
Db 111 GATTGATGTGATATCTCCACTGACGTAAAGGATGACCA 149

Search completed: May 11, 2003, 04:54:21
Job time: 1272.94 secs

